Text

Description automatically generated

**“Sentiment-Driven Topic Analysis in ISIS Tweets: Exploring Sentiment-Topic Pairs through Semi-Supervised and Unsupervised LDA”**

***Group 2***

Lorielle Angelei Villegas Malveda: [lav.malveda.2023@mitb.smu.edu.sg](mailto:lav.malveda.2023@mitb.smu.edu.sg)

Huang Shaofei: [sf.huang.2023@engd.smu.edu.sg](mailto:sf.huang.2023@engd.smu.edu.sg)

Goh Cheng En: [chengen.goh.2023@mitb.smu.edu.sg](mailto:chengen.goh.2023@mitb.smu.edu.sg)

Low Wei Jun: [weijun.low.2023@mitb.smu.edu.sg](mailto:weijun.low.2023@mitb.smu.edu.sg)

Sarthak Jagdish Nagapurkar: [sarthakj.n.2023@mitb.smu.edu.sg](mailto:sarthakj.n.2023@mitb.smu.edu.sg)

Liang Xiuhao: [xh.liang.2023@mitb.smu.edu.sg](mailto:xh.liang.2023@mitb.smu.edu.sg)

***9 November 2024***

# **1.0 Introduction**

The rise of social media has provided a platform for free expression of views on both positive and negative communication. One of the more troubling aspects is how extremist groups, such as ISIS, leverage platforms like X (formerly known as Twitter) to spread their propaganda, recruit members and coordinate activities. Analysing these online activities is of growing importance for governments, intelligence agencies, law enforcement and social media platform moderators to detect and counter these harmful narratives in a timely fashion.

## 1.1 Use Case Scenario and Dataset

We will analyse a dataset of ISIS Tweets from Kaggle that contains tweets related to ISIS activity and narratives from January 2015 to January 2016. The dataset offers a comprehensive collection of tweets that are crucial for understanding how ISIS-affiliated individuals and groups use social media to spread their ideology, recruit members and coordination of movements and activities. Given the large volume of data and the sensitive nature of the content, it is essential to apply text analytics to detect trends, sentiment shifts and potential threats in real-time.

# **2.0 Project Objectives**

## 2.1 Analytics Tasks

The main objective of this project is to perform sentiment analysis on the ISIS-related tweets dataset by classifying the tweets into three categories: positive, negative, and neutral before they undergo further analysis to determine if they fall into more specific topics, such as propaganda, hate, and planned attacks etc. This deeper level of topic modelling will help uncover their intent, especially those promoting extremist narratives or encouraging harmful activities.

## 2.2 Usefulness of the Analysis and its Potential Users

Analysing sentiment in ISIS-related tweets is useful for several reasons:

1. **Early Detection of Extremist Messaging**: Identifying emotionally charged tweets that signal propaganda, recruitment efforts, or attack coordination allow authorities to intervene early.
2. **Improved Content Moderation**: Social media platforms can benefit from sentiment analysis to develop more nuanced content moderation strategies that go beyond keyword filtering and focus on context and intent.
3. **Strategic Countermeasures**: Law enforcement, intelligence and counter-terrorism agencies can use the analysis to detect emerging trends in online radicalization to refine their strategies and take pre-emptive measures.

# **3.0 Methodology**

We performed sentiment analysis to label tweets as positive, negative or neutral using various language models namely Valence Aware Dictionary and Sentiment Reasoner (VADER) and AFINN, which are lexicon and rule-based sentiment analysis models, as well as BERTweet, a pre-trained transformer model. We then selected a suitable language model by comparing the tweet categorisation based on the respective sentiment labels. As the original dataset does not include sentiment labels, there is no ground truth to compare the results against. As such, we will base our assessment on intuition by sampling and examining tweet contents from the sentiment-labelled tweets. Once the tweets are categorised, we will apply topic modelling to each sentiment group to extract key topics.

**Analysis Steps:**

1. Data Preprocessing: Data cleaning and normalisation, exploratory data analysis (EDA), feature extraction using word embedding
2. Sentiment Analysis: Label tweets using VADER, AFINN and BERTweet into based on sentiments (Positive, Negative, Neutral), then performing additional EDA on respective results
3. Topic Modelling: (1) Semi-supervised learning through manually labelling a portion of the data with pre-defined seed topics and words through TF-IDF in LDA. (2) Apply LDA to discover topics within each sentiment group.
4. Analysis: Identify accounts and tweets that express pro-ISIS sentiments based on sentiment topic pair.

# **4.0 Solution Details**

## 4.1 Sentiment Analysis Using VADER, AFINN and BERTweet Concurrently

We applied VADER, which is optimized for short texts like tweets. VADER assigns a sentiment score ranging from -1 (most negative) to +1 (most positive), with a compound score classifying tweets as positive, neutral, or negative. AFINN assigns a sentiment score to words ranging from -5 (most negative) to +5 (most positive). The total sentiment of a tweet is calculated by summing the scores of all the words in the text. AFINN is particularly suited for quick sentiment analysis and is straightforward in its application, making it useful for general-purpose text sentiment classification. BERTweet uses deep learning to grasp the context and meaning of words in a tweet. This model fine-tunes on a large corpus of Twitter data, making it effective in capturing complex language, including slang, acronyms and nuanced expressions.

## 4.2 Advantages and Limitations Between VADER, AFINN and BERTweet

This table summarizes the advantages and limitations of the different language models:

|  |  |  |
| --- | --- | --- |
| Model | **ADVANTAGES** | **LIMITATIONS** |
| **VADER** | 1. Pre-trained for social media data.  2, Effective at categorizing nuanced text.  3. Handles slang and emoticons well. | 1. Difficulty in detecting sarcasm.  2. Hidden sentiment can be misclassified.  3. Scores may be skewed by capitalization or punctuation. |
| **AFINN** | 1. Simple and computationally efficient.  2. Works well with direct emotions  Easily adaptable to different languages | 1. Limited vocabulary may cause misclassification  2. Lacks negation handling, affecting sentiment accuracy  3. Struggles with context, irony, and sarcasm |
| **BERTweet** | 1. Context-aware, handling complex language well  2. Generalizes across various tweet structures  3. Fine-tuned for Twitter data, ideal for social media analysis | 1. Computationally intensive compared to lexicon-based approaches.  2. Less suited for real-time analysis in resource-limited settings  3. Requires substantial data and hardware for best performance |

## 4.4 Topic Modelling Within Sentiment Categories

Once the tweets were classified by sentiments, we applied Latent Dirichlet Allocation (LDA) to uncover key topics within each sentiment group, using both semi-supervised LDA, where predefined labels guide topic generation and unsupervised LDA, where topics are discovered solely from the data. After identifying the key topics in positive sentiment tweets, we will analyse the accounts that frequently engage with these topics. This will help pinpoint potential ISIS sympathizers or propaganda sources.

**Advantages**:

LDA uncovers hidden structures in text by clustering words into coherent topics. It helps identify prominent topics in positive sentiment tweets, potentially revealing pro-ISIS narratives.

**Limitations**:

* Short text like tweets might result in weak topic coherence.
* Requires careful tuning of the number of topics to ensure relevant insights.
* Choosing appropriate labels for semi-supervised LDA can be complex.

## 4.5 Semi-supervised LDA

### 4.5.1 Rationale for Selecting Seed Topics

For this project, we focused on three primary seed topics: **propaganda**, **recruitment**, and **radicalization**. These topics were selected for several key reasons:

* **The primary purposes for which ISIS uses online platforms**: According to research on ISIS and extremist online activities, these three topics—propaganda, recruitment, and radicalization—are the primary ways ISIS spreads its influence and operational objectives on social media.
* **The seed word list was generated from research that uses ISIS and other extremism texts for their labels**: This approach ensures that our seed words are highly relevant and reflect real-world extremist communications. The words we used are based on existing academic research that focuses on extremist text data and language patterns associated with groups like ISIS.
* **Some seed words extracted from popular ISIS publications (e.g. Dabiq)**: In addition to online posts, ISIS regularly uses similar language patterns in their offline materials, such as publications like *Dabiq*. These materials often contain words that are mirrored and used in their social media campaigns, making it ideal for identifying online extremist messages.

# **5.0 Experiments**

## 5.1 Dataset Description

The dataset consists of 17,410 entries with the following columns:

1. name: name associated with the Twitter account
2. username: twitter handle
3. description: bio of the Twitter account
4. location: location of the user
5. followers: number of followers the account has
6. number statuses: number of tweets or statuses the account has posted
7. time: date/timestamp when the tweet was posted
8. tweets: content of the tweets

Exploratory Data Analysis (EDA) is performed on the original tweets to analyse and identify issues that need to be addressed during data pre-processing. First, we calculate the number of words in each tweet in the dataset to see if there are outliers in the distribution.



We observe that tweets with140 words have the highest frequency in the dataset. There were no outliers i.e. multiple peaks in the distribution. Also, there were no tweets of 160 words and above – this can be attributed to the word limit of X (Twitter) postings.

Next, we constructed a word cloud based on the dataset to detect if there were any unnecessary words or characters.



We can identify from the word cloud several unnecessary words e.g. ”https”, ”RT” (which stands for ”re-tweet”) etc. These were removed in the next stage.

## 5.2 Data Preprocessing Techniques

Data Cleaning using Regex:

* **Translation-related text:** As some of the original tweets are in Arabic, the tweets that begin with “ENGLISH TRANSLATIONS” will have these words removed.
* **Non-alphabetic characters:** Strip all non-alphabetic characters except the “#” symbol used for hashtags.
* **Retweet indicators, handles, and URLs**: Remove "RT @", “http”, “https”, and “www”.
* **Unnecessary text:** Remove “amp” which stands for ampersand, used as a string literal to represent “&” in the dataset.

Hashtag and words separation:

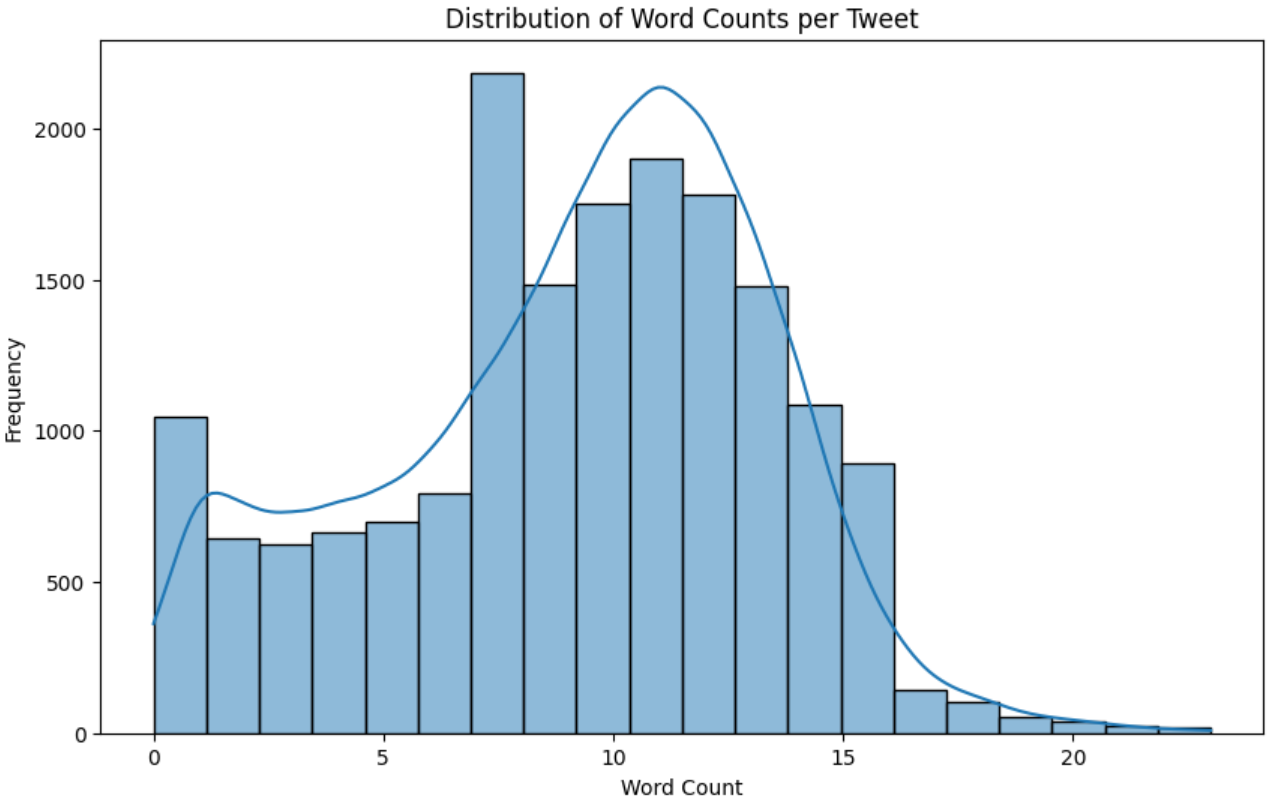
* Separate hashtags from other words and capture the remaining words within the tweets.

Data Normalisation through Stemming and Lemmatisation:

* Option 1: Remove stop words and stem words using Porter stemmer
* Option 2: Remove stop words and lemmatize words

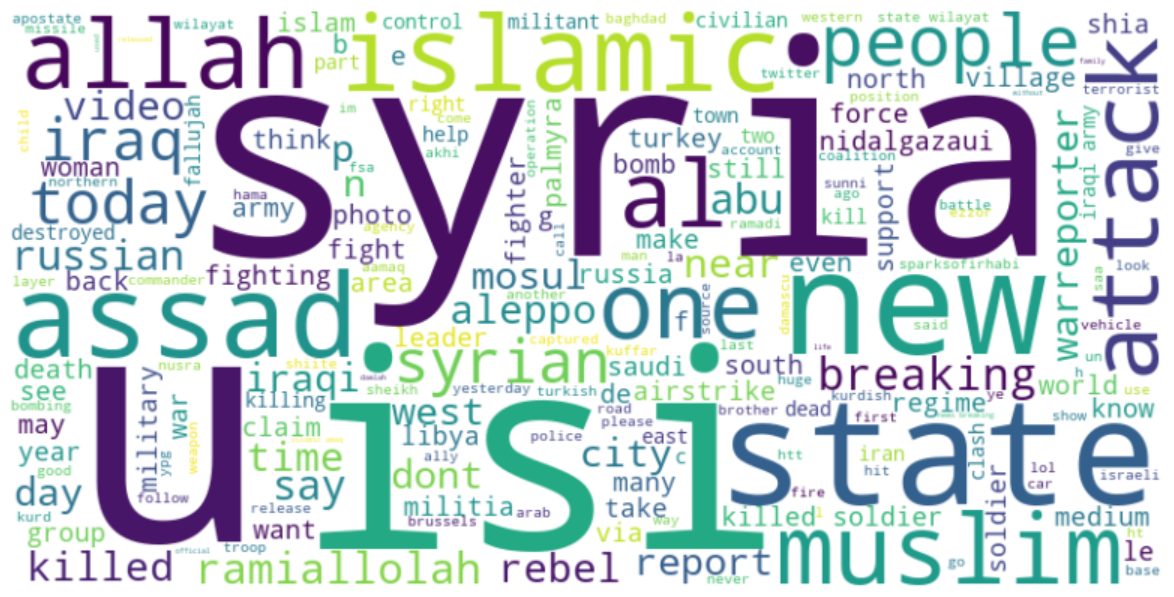
We experimented with the two options and the detailed results can be found in **Appendix 1**. The results show that Option 2 is more suitable for the dataset, to preserve the meaning of words in the tweet contents – an important requisite for effective sentiment analysis.

After pre-processing, another round of EDA is performed on the processed tweets. First, we calculate the number of words in each tweet in the dataset to see if there are outliers in the distribution.



We observe that the processed tweets display a somewhat normalised distribution.

Next, we construct a word cloud based on the processed dataset to confirm that unnecessary words or characters have been removed.

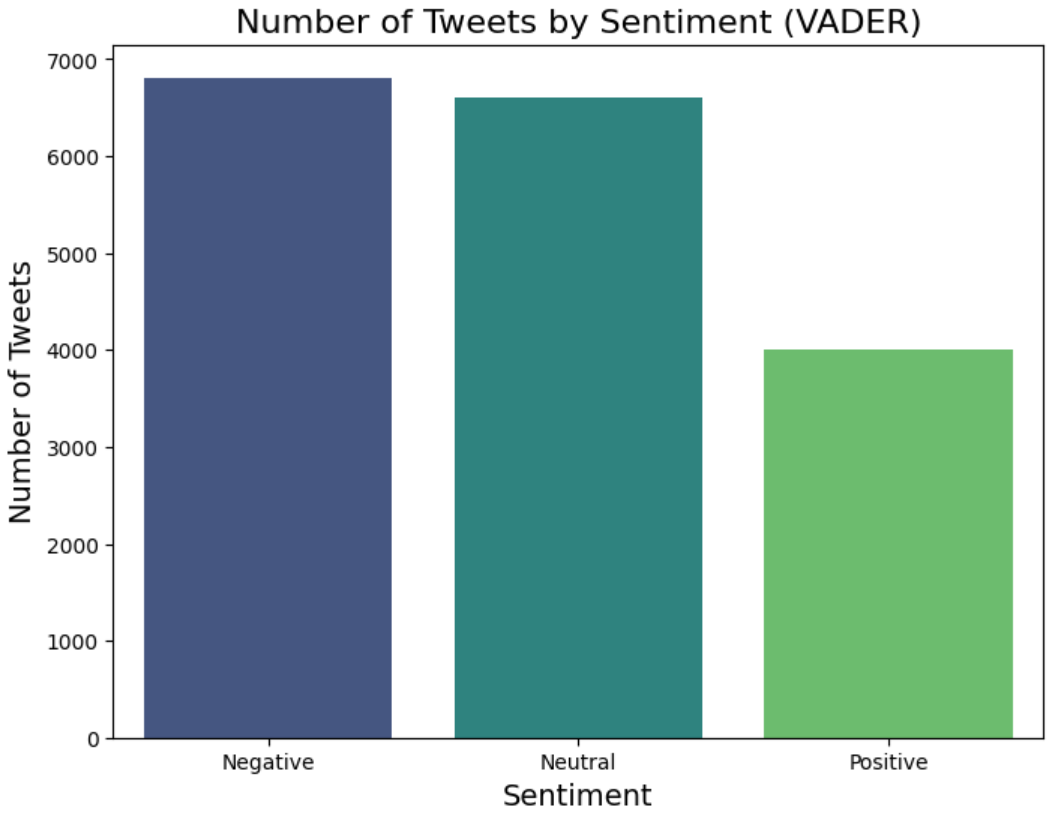


## 5.3 Data Annotation and Labelling

We experimented with three different methods of sentiment analysis, namely: VADER, AFINN and BERTweet (based on BERT). Since ground-truth sentiment labels isn’t used in the dataset, the models are evaluated intuitively through sentiment-labelled data and corresponding text.

### 5.3.1 Sentiment Analysis using VADER, AFINN & BERTweet

The results of the sentiment analysis by VADER, AFINN and BERTweet were as follows:

A graph of a number of tweets

Description automatically generatedA graph of a number of tweets

Description automatically generated

The nuanced sentiment detection with a higher volume of neutral tweets and reduced positivity bias with a smaller proportion of positive tweets compared to the other 2 showed that BERTweet was more superior. This is in line with how BERTweet as a transformer-based model, which was pre-trained on Twitter data, leverages context more effectively than lexicon-based models like VADAR and AFINN.

5.4 Data Partition Strategy

There is no data partitioning of the dataset as the above-mentioned methods for sentiment analysis do not require training, validation and evaluation of models. Since the BERTweet is a pre-trained transformer, and we are not training and validating the model further in our analysis.

# 5.5 Data Preprocessing for Topic Modelling

It was observed that words from foreign languages, like French, contributed noise to the topic modeling results. Additionally, commonly used informal English words in tweets, along with special characters and spelling errors, added further noise. The data was pre-processed to remove frequently occurring foreign words and to handle spelling errors.

Preprocessing involved the following steps:

* Removal of top French foreign words.
* Removal of commonly used words in informal English.
* Correcting words with special characters and spelling errors.

# 5.6 Unsupervised LDA using Gensim

## 5.6.1 Implementing unsupervised LDA using Gensim

The Gensim package is a Python library designed for unsupervised topic modeling and document similarity analysis, with efficient implementations for tasks like LDA.To perform unsupervised LDA for all three sentiment tweets (Positive, Negative and, neutral) in Gensim, the following steps were followed

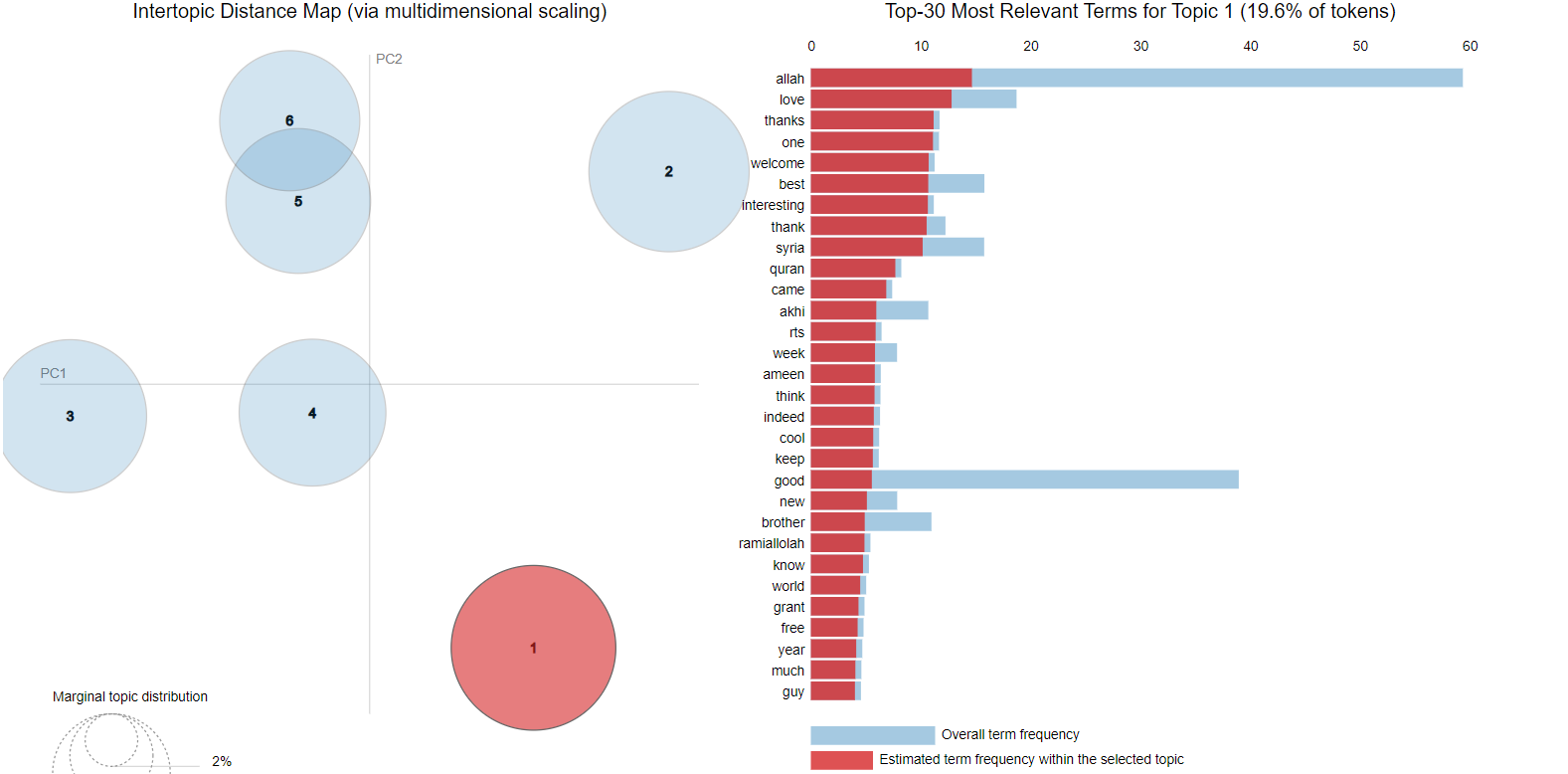
* After preprocessing text data, a dictionary and bag-of-words corpus was created.
* LDAModel was used to extract relevant topics from the corpus.

## 5.6.2 Visualizing topic modelling results through pyLDAvis

The **pyLDAvis** library is a Python tool for interactive visualization of topic modeling results, where each topic is displayed as a circle on a 2D plane, with size indicating topic prevalence and distance indicating topic similarity.

## 5.6.3 Interpreting topic modelling results

Let’s interpret the results for topic modelling for Positive sentiment.



Key takeaways:

* Prominent words for topic 1: *allah, love, thanks, one, welcome.*
* The number of prominent topics can be estimated at 5-6.
* Topic 1 is the most prominent in the positive sentiment tweets consuming 19.6% of tokens.
* Slight overlap of tokens between topics 5 & 6.

# 5.7 Semi-Supervised LDA using Gensim

Results of unsupervised topic modeling tend to be noisy and may miss relevant topics. Hence it was essential to perform semi-supervised LDA. Semi-supervised LDA is better than unsupervised LDA because it integrates labeled data to guide topic formation. It helps produce more relevant and interpretable topics aligned with specific domains or categories. We proceed with identifying seed topics and seed words across the three sentiments.

## 5.7.1 Identifying Seed topics across sentiments

We broadly had the following research aims to guide topic modeling for all three sentiments.

* **Early Detection of Extremist Messaging** – Differentiating between impeding attacks and recruitment instructions.
* **Improved Content Moderation** – Differentiating radicalisation from other legitimate religious texts.
* **Strategic Countermeasures** – Identify terrorist attack targets, counter-recruitment, counter-radicalization, and labelling of at-risk persons based on tweet follows or RT.

The themes corresponding to these three aims were identified as follows.

* Spreading Propaganda [Propaganda]
* Attracting youths through recruitment messages [Recruitment]
* Radical Change in perception towards an individual or community [Radicalisation].

The seed topics and seed words within the sentiment tweets corresponding to these three themes were identified for all three sentiments (Positive, Negative & Neutral).

## 5.7.2 Performing Semi-Supervised LDA using Gensim

We can follow the below steps to perform semi-supervised LDA for all three sentiment-based tweet using Gensim:

* Use the existing dictionary and corpus created for unsupervised LDA.
* Use TF-IDF to come up with the seed words and merge them with the seed topics. Refer to **Appendix 2**.
* Incorporate identified seed topics and seed words in the creation of the LDA model.
* LDA model was used to extract topics from the corpus.
* Visualize results of semi-supervised LDA using pyLDAvis. Inspect the impact the seed words and seed topics had in extraction of topics.

## 5.7.3 Interpreting results of Semi-Supervised LDA using Gensim

Let’s interpret the results for semi-supervised LDA for Negative sentiment.

A graph with a red circle and text

Description automatically generated

The key take away is that Topic 1 is the most prominent in the negative sentiment tweets consuming 50.7% of tokens. This topic aligns with the manually chosen seed topic: **Terrorist Operations and Violence**. Next, manually identified seed words for the topic Terrorist Operations and Violence: *killed, Syria, and army* are the top relevant terms for topic 1 in terms of frequency indicating the seed words and seed topics have significantly biased the topic modeling results.

# 5.8 Topic Modeling using GPT 4.0

As a final step, we uploaded the three sentiment-labeled tweet datasets to GPT-4.0, prompting it to identify key topics, associated themes, and corresponding seed words for each of them. The goal was to assess whether the topics identified by unsupervised and semi-supervised LDA aligned with those generated by GPT-4.0.

# 6.0 **Results and Analyses**

The tables below present the perplexity and coherence scores, for each sentiment.

**Unsupervised LDA**

|  |  |  |
| --- | --- | --- |
| Sentiment | Perplexity | Coherence Score |
| Positive | -8.00 | 0.60 |
| Neutral | -9.32 | 0.36 |
| Negative | -8.45 | 0.38 |

**Semi-Supervised LDA**

|  |  |  |
| --- | --- | --- |
|  | Perplexity | Coherence Score |
| Positive | 3.21 | 0.56 |
| Neutral | -5.26 | 0.42 |
| Negative | -5.17 | 0.35 |

In the context of topic modeling, a lower perplexity score indicates a better fit to the data, meaning the topics generated by the model more accurately represent the text data. On the other hand, higher coherence scores suggest that the topics make more sense to human readers because the words within each topic are more closely related.

***PERPLEXITY SCORES***

In the Unsupervised LDA model, negative perplexity scores for all sentiments indicate a relatively good fit, with the "Neutral" sentiment having a better score (-9.32), than the others.

In the Semi-Supervised LDA model, the perplexity scores are higher and positive for the "Positive" sentiment, but lower for the "Neutral" and "Negative" sentiments, which suggests that the model struggles more with fitting the data compared to the Unsupervised LDA model.

***COHERENCE SCORES***

For both models, the "Positive" sentiment shows the highest coherence scores, indicating that the positive topics are the most interpretable and cohesive. In the Unsupervised LDA model, the "Positive" coherence score is 0.60, higher than both "Neutral" and "Negative."

The Semi-Supervised LDA model has a slightly lower coherence score for "Positive" at 0.56 but performs similarly to the Unsupervised LDA model. The supervised learning model may improve structure in topic relevance, though perhaps at some cost to coherence.

Overall, the Unsupervised LDA model performs better in terms of perplexity, while both models show comparable coherence scores.

# 6.1 Comparison of Topics across three methods

(Refer to **Appendix 3, 4 & 5** for details of each topic)

## 6.1.1 Sentiment-Topic Analysis

When we compare the topics within each sentiment, we observe that positive sentiment topics tend to converge on religious expressions, with terms suggesting non-violent but supportive messaging that promotes group identity and recruitment. Even for seemingly negative words like martyrdom, the topic suggested by LLM resulted in pairing martyrdom with heroism. For positive sentiments and topic pair, we may conclude that the sentiment-topic pairs identified are largely for propaganda and recruitment purposes.

For neutral sentiment topics, they tend to converge on practical updates, like statements and updates on group activities. These narratives are factual in nature but nonetheless still include religious and ideological elements, positioning the group’s activities within a broader geopolitical context. In a practical sense, these topics may not invoke strong emotions from the audience but may appeal to individuals who prefer factual message in support of ISIS.

For negative topics, across the different methods used, the convergence is along the line of military conflict, civilian impact with emotionally charged language and word patterns. The topics revolve around radicalization and fostering hostility. The focus on casualties and opposition creates a victim-persecutor dynamic, potentially appealing to individuals sympathetic to militant causes.

## 6.2.2 Business use-case analysis

For positive sentiment, content moderators can identify subtle propaganda and monitor recruitment messaging by tracking religious and community terms tied to ideological expressions. Social media platforms might focus on clusters that frequently use these terms, flagging content that indirectly supports extremist ideologies through non-violent messaging.

For neutral sentiment, early detection of extremist messaging can help intelligence agencies track territorial control and group activities. Understanding these narratives helps in assessing which geographic areas or groups are of interest to extremists. Also, social media platforms can improve content moderation by distinguishing legitimate news from covert propaganda.

For negative sentiment, counter-terrorism agencies and social media platforms can use this analysis to monitor narratives involving military conflicts and anti-Western sentiments. It could assist to track radicalization hotspots and recruitment points. Humanitarian organizations can use the insights to assess areas with increased civilian impact and allocate resources effectively.

## 6.2.3 Practical Applications

After analyzing the topics using three different methods, we discovered that besides our use cases that we have mentioned in section 2.2, there are more other specific applications when using the sentiment-topic pairing.

|  |  |  |  |
| --- | --- | --- | --- |
| **Business Use Case** | **Positive** | **Neutral** | **Negative** |
| Content Moderation | ✔ | ✔ | ✔ |
| Early Detection of Extremist Messaging | ✔ | ✔ | ✔ |
| Tracking Recruitment Indicators | ✔ |  | ✔ |
| Monitoring Morale and Ideological Support | ✔ | ✔ |  |
| Real-time Crisis Monitoring for Humanitarian Aid |  | ✔ | ✔ |
| Geopolitical Analysis and Risk Assessment |  | ✔ | ✔ |
| Identifying Key Influencers and Networks | ✔ | ✔ | ✔ |
| Strategic Countermeasures for Intelligence Agencies | ✔ | ✔ | ✔ |

# **7.0 Discussion and Gap Analysis**

One significant limitation was the dataset’s lack of labeled sentiment categories, which made it difficult to evaluate sentiment analysis models effectively. Without “ground-truth” sentiment labels, we lacked a reliable metric to validate the sentiment predictions. Furthermore, the dataset's small size and its inherent bias (leaning heavily toward pro-ISIS sentiment) caused an imbalance, skewing the sentiment distribution and reducing the representativeness of the data.

# **8.0 Future Work**

In our analysis, we did not consider French words that remained in the corpus. These words were in the corpus as the French language used alphabets, and were excluded from the Arabic-English translation of the original tweets. What could be done in future studies is to properly identify the different languages and be deliberate about the translation. For this analysis, we only used ISIS supporter tweets. In future works, study can be done to compare pro-ISIS against anti-ISIS sentiment-topic pair to have a clear differentiation between the two groups of users. This will help content moderators to get to the correct accounts as the central nodes for information dissemination.

# **9.0 Project Experience/Reflections**

## Refer to **Appendix 6**

**References**

Please use APA style. <https://pitt.libguides.com/c.php?g=12108&p=64730>

1. Hugging Face. (n.d.-a). *BERTweet documentation*. Retrieved November 5, 2024, from <https://huggingface.co/docs/transformers/en/model_doc/bertweet>

2. Hugging Face. (n.d.-b). *BERTweet-base sentiment analysis*. Retrieved November 5, 2024, from <https://huggingface.co/finiteautomata/bertweet-base-sentiment-analysis>

3. Hutto, C. J., & Gilbert, E. E. (2014). *VADER: A parsimonious rule-based model for sentiment analysis of social media text*. Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM-14), Ann Arbor, MI, June 2014.

4. Nguyen, D. Q., Vu, T., & Nguyen, A. T. (2020). BERTweet: A pre-trained language model for English tweets. *arXiv preprint* arXiv:2005.10200. <https://arxiv.org/abs/2005.10200>

5. PyPI. (n.d.). *Afinn*. Retrieved November 5, 2024, from <https://pypi.org/project/afinn/>

# APPENDICES

## Appendix 1

**Option 1 – Remove stop words and stem words using Porter stemmer**

0 [sheikh, muham, abu, al, maqdisi, syria, truth...

1 [sheikh, integr, easi, fatih, al, jawlani, peo...

2 [sheikh, meet, fatih, first, al, jawlani, ha, ...

3 [leader, sheikh, nasir, al, aqap, ha, promis, ...

4 [sheikh, although, dislik, respons, aqap, disb...

Name: wordlist, dtype: object

count 17410

unique 16560

top []

freq 280

Name: wordlist, dtype: object

**Option 2 – Remove stop words and lemmatize words**

0 [sheikh, muhammed, abu, al, maqdisi, syria, tr...

1 [sheikh, sacrifice, fatih, al, jawlani, people...

2 [sheikh, fatih, first, al, jawlani, ha, audio,...

3 [leader, sheikh, nasir, promise, al, aqap, ha,...

4 [baghdadis, sheikh, response, although, dislik...

Name: wordlist, dtype: object

count 17410

unique 16560

top []

freq 280

Name: wordlist, dtype: object

## Appendix 2

Positive Sentiment-Topics with Seed Words

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **S/N** | **Business Use Case** | **Topics Modelling (Positive)** | **Seed Word 1** | **Seed Word 2** | **Seed Word 3** | **Seed Word 4** | **Seed Word 5** |
| 1 | Community Engagement: Monitoring positive community interactions related to counter-extremism efforts. | Community Engagement | love | good | great | best | amazing |
| 2 | Positive Messaging:  Identifying supportive and uplifting messages that counter extremist narratives | Uplifting Messages and Support | victory | alhamdulillah | interesting | support | one |
| 3 | Advocacy:  Promoting supportive narratives that help counter negative perceptions | Positive Advocacy and Representation | help | beautiful | better | thanks | day |
| 4 | Empowerment: Identifying messages that empower individuals and communities | Resist Extremist Idealogies | back | new | video | iraq | state |

Neutral Sentiment-Topics with Seed Words

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **S/N** | **Business Use Case** | **Topics Modelling (Positive)** | **Seed Word 1** | **Seed Word 2** | **Seed Word 3** | **Seed Word 4** | **Seed Word 5** |
| 1 | Data Monitoring: Tracking general sentiment in discussions aids in understanding public perception without bias. | Sentiment Analysis | isis | al | syria | allah | us |
| 2 | Information Dissemination: Monitoring neutral discussions helps inform stakeholders without inciting emotion. | Neutral Discussions | attack | islamic | city | video | near |
| 3 | Educational Outreach: Promoting educational resources can lead to better informed communities. | Knowledge Sharing | city | new | ramiallolah | one | islam |
| 4 | Public Awareness: Raising Awareness about issues without inciting strong emotions or biasness | Propaganda | attack | fighters | rebels | forces | state |

Negative Sentiment-Topics with Seed Words

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **S/N** | **Business Use Case** | **Topics Modelling (Positive)** | **Seed Word 1** | **Seed Word 2** | **Seed Word 3** | **Seed Word 4** | **Seed Word 5** |
| 1 | **Threat Detection and Risk Assessment:** Identifying tweets mentioning violent actions (e.g., bombings or attacks) helps predict potential attacks. | Terrorist Operations and Violence | *killed* | syria | army | bombing | fighters |
| 2 | **Platform Moderation:** Detecting organizational messages helps identify accounts spreading extremist influence. | ISIS Leadership and Affiliates | *isis* | al | assad | forces | syrian |
| 3 | **Strategic Communication:** Identifying anti-Western themes helps governments design counter-messages to reduce ISIS's influence over local populations. | Anti-Western Sentiment or Foreign Intervention | *us* | west | aleppo | iraqi | policy |
| 4 | **Radicalization Prevention:** Monitoring religiously framed tweets helps detect potential radicalization trends among individuals. | Religious Justifications or Propaganda | *allah* | jihad | faith | martyr | sharia |
| 5 | **Threat Detection:** Tracking discussions about territorial changes or control helps predict new conflict zones or areas at risk of falling under ISIS influence. | Geopolitical Impact on Cities or Areas under ISIS Influence | *city* | state | territory | region | districts |
| 6 | **Counter-Recruitment Campaigns:** Recognizing calls for unity or sacrifice enables interventions to dissuade individuals from joining ISIS. | Ideological Messaging or Recruitment Themes | *one* | unity | brotherhood | sacrifice | loyalty |

## Appendix 3

**Table A3.1: Positive Sentiment**

|  |  |  |  |
| --- | --- | --- | --- |
| **Topic No.** | **Unsupervised LDA** | **Semi-Supervised LDA** | **LLM (ChatGPT)** |
| 1 | Religious and Spiritual Expressions | Faith and Resilience in Challenging Times | Religious Unity and Moral Justification |
| 2 | Positivity and Brotherhood | Social Media Engagement and Global Events | State-Building and Ideological Goals |
| 3 | Hope and Progress | Expressions of Faith, Support, and Community Connection | Military Power and Regional Influence |
| 4 | Support and Celebration | Expressions of Positivity, Resilience, and Support | Martyrdom and Heroism |
| 5 | Community and Pride | Faith, Respect, and Understanding | Community Engagement and Recruitment Appeals |
| 6 | Enthusiasm and Leadership |  |  |

**Table A3.2: Neutral Sentiment**

|  |  |  |  |
| --- | --- | --- | --- |
| **Topic No.** | **Unsupervised LDA** | **Semi-Supervised LDA** | **LLM (ChatGPT)** |
| 1 | Military Activities | Support and Group Dynamics | Conflict Zone Updates |
| 2 | Religious and Ideological Support | Military and territorial Updates | Cultural and Religious Discourse |
| 3 | Military Operations and Governance | Religious and Social | *Political Developments in Syria* |
| 4 | Territorial Control and Group Activities |  | Leadership Statements and Public Announcements |
| 5 |  |  | *Community Events and Observances* |

**Table A3.3: Negative Sentiment**

|  |  |  |  |
| --- | --- | --- | --- |
| **Topic No.** | **Unsupervised LDA** | **Semi-Supervised LDA** | **LLM (ChatGPT)** |
| 1 | Conflict and Military Operations | Warfare and Combat Casualties | Casualties and Military Confrontations |
| 2 | ISIS and Civilian Impact | Terrorism Impact on Civilians | Criticism of Opposing Groups |
| 3 | Regime, Rebels and Fighting | Ideological Conflicts in Syria | Calls for Solidarity against Adversaries |
| 4 | Radical Ideology and Western Opposition |  | Breaking News on Regional Violence |
| 5 |  |  | Martyrdom and Revenge Narratives |

## Appendix 4

**Unsupervised LDA output**

# POSITIVE

## TOPIC 1

**Religious and Spiritual Expressions**

This topic contains words like **Allah, alhamdulillah (praise to God), ameen (amen), prophet, reward, and Qur'an**, which are strongly associated with expressions of faith, praise, and religious devotion. The inclusion of **khilafah, Abu, and operation** uggests ties to religious figures or leaders. However, the overall sentiment seems to revolve around **spirituality, blessings, and faith**.

**Possible Label**: **Faith and Religious Gratitude**

*allah, love, thanks, one, welcome, best, interesting, thank, syria, quran, came, akhi, rts, week, ameen, think, indeed, cool, keep, good, new, brother, ramiallolah, know, world, grant, free, year, much*

**Business Use Case and Practical Application:**

In this sentiment-topic tag focused on positive-faith, the language emphasizes spiritual devotion and religious community, which appeals to audiences seeking a sense of belonging rooted in faith. Terms such as ameen and alhamdulillah foster a sense of shared reverence, which can subtly cultivate sympathizers through expressions of religious solidarity. Although non-violent, this messaging can indirectly support ISIS by promoting their religious ideals, creating an ideological gateway for sympathizers to identify with the group. For content moderation, identifying accounts and clusters of religious expression that frequently mention khilafah and specific religious leaders can highlight early ideological alignments that may lead to more radical expressions.

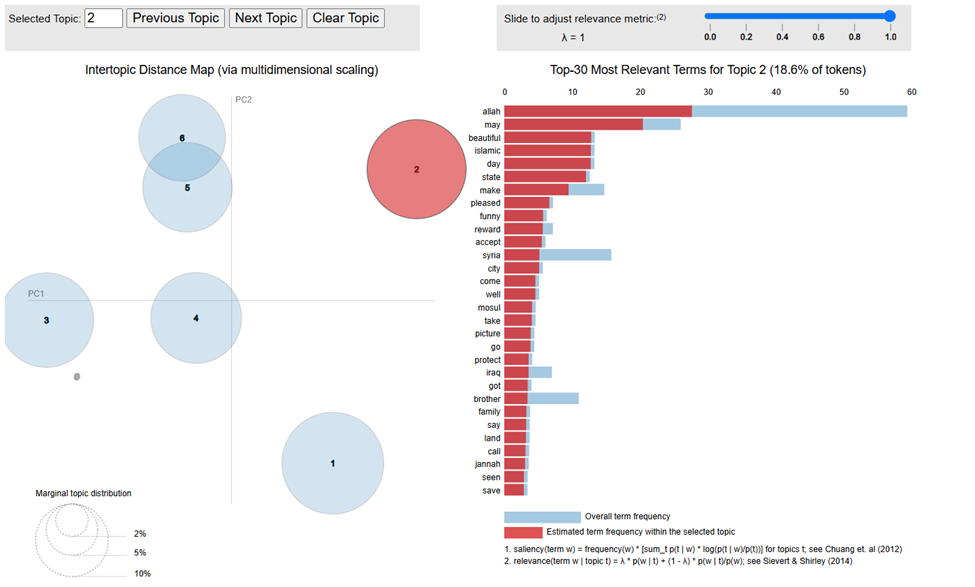
## TOPIC 2

**Positivity and Brotherhood**

This topic emphasizes **positivity**, with words like **beautiful, amazing, welcome, brother, and love**. It seems to capture expressions of **community, care, and well-being**, likely reflecting messages of encouragement or solidarity within the Muslim community. Terms like **grant, quran, and life** suggest religious undertones and focuses on positive feelings and community connection.

**Possible Label**: **Positivity and Brotherhood:**

*allah, may, beautiful, islamic, day, state, make, pleased, funny, reward, accept, syria, city, come, well, mosul, take, picture, go, protect, iraq, got, brother, family, say, land, call, jannah, seen, save*



***Business Use Case and Practical Application:***

*The sentiment-topic pair positive-brotherhood uses inclusive language (brother, love, welcome) to build a sense of community within the Muslim identity. By promoting messages of unity and solidarity, these expressions serve to unify supporters under a shared cultural and religious bond, establishing a supportive network. This sense of kinship, while seemingly benign, is a potential gateway for ISIS sympathizers to deepen connections with the ideology subtly. Content moderators should monitor clusters that emphasize brotherhood within a pro-ISIS context, as these may be active in mobilizing new sympathizers or reinforcing morale. Tracking how terms of unity are used alongside identifiers like Syria or Iraq could reveal recruitment pipelines or logistical nodes that align positive messaging with extremist goals.*

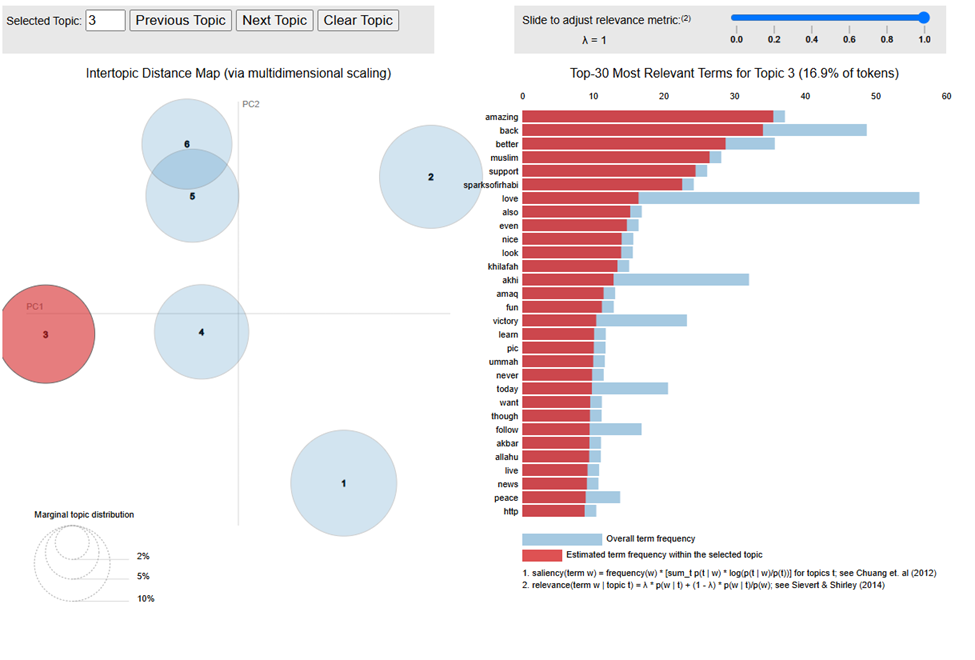
## TOPIC 3

**Hope and Progress**

This topic includes words like **good, peace, bless, blessing, military, and fighting**, indicating themes of **optimism and possibly victory**. It could relate to positive developments, as the words **new, force, learn**, and **blessing** suggest growth or progress. The term **Islamic State** might reflect the context, but the focus is more on **hopeful outcomes**.

**Possible Label**: **Hope and Positive Change**

*amazing, back, better, muslim, support, sparksofirhabi, love, also, even, nice, look, khilafah, akhi, amaq, fun, victory, learn, pic, ummah, never, today, want, though, follow, akbar, allahu, live, news, peace*



***Business Use Case and Practical Application:***

*In this sentiment-topic tag focused on positive-hope, words of optimism (blessing, good, peace) serve to paint a future aligned with the ideals of the Islamic State. By portraying growth and progress, these messages can uplift morale, inspire action, and cultivate a loyal supporter base. The combination of hopeful language and military references (fighting, force) serves to connect personal faith and positivity with ISIS’s objectives, making the ideology more palatable to sympathizers. Content moderators can use this hopeful rhetoric to track influencers and posts that may serve as soft recruitment tools, aiming to convert neutral or mildly supportive individuals into active participants by associating hope with the group’s cause.*

## TOPIC 4

**Support and Celebration**

This topic revolves around words like **thanks, support, love, happy, celebrate, and huge**, suggesting celebrations or expressions of gratitude. The word **fallujah** may indicate that this celebration is tied to some **military success or victory**, while the other terms suggest a **sense of joy and community support**.

**Possible Label**: **Celebration and Gratitude**

*good, great, see, allah, video, like, aleppo, unclesamcoco, sheikh, victory, way, lol, work, name, subhanallah, fallujah, must, understand, strong, operation, soldier, de, brave, many, seems, eastern, respect, si, watch, bless*



***Business Use Case and Practical Application:***

*In this sentiment-topic tag* ***positive-support****, we find many words (thanks, love, happy, cool, celebrate, brother, interesting, much, huge, yes, funny, indeed, yeah) that are expressing positive emotions, celebrations, camaraderie, or humor, often used in supportive contexts. Moreover, words such as* ***brother*** *reflects kinship, typically among members of extremist groups, indicating solidarity and recruitment messaging. This is coupled with action-oriented words (make, get, go, call, support, saw, way, back) suggesting recruitment and mobilisation of fighters to respond to the call. To round up the narrative, the use of words such as syria, fallujah, allahu, akbar, strengthen the religious narrative to support ISIS in the fight against the west. All in all, this topic-sentiment pair suggest support and celebratory tweets that are skewed towards* ***propaganda and recruitment effort****. These tweets do not depict violence, and may not be targeting at recruiting fighters. Instead, it may be targeting ISIS sympathisers around the world, to maintain morale, cultivate global support, and normalise ISIS ideology among sympathizers. These tweets likely aim to create a sense of belonging and shared identity, fostering an emotional connection that strengthens allegiance to the cause. Through non-violent messaging, ISIS can subtly build networks of supporters who may amplify its propaganda or provide logistical aid, even if they are not directly involved in combat. This type of messaging helps to sustain the group’s ideological footprint and ensures continuity in influence across dispersed sympathizers. With words such as “nidalgazaui” identified in the list of words in this topic which looks like a twitter account,* ***content moderators can zoom into such accounts*** *that likely push for positive-sentiment ISIS propaganda, which may provide further linked accounts that provides logistical support in the whole ISIS network.*

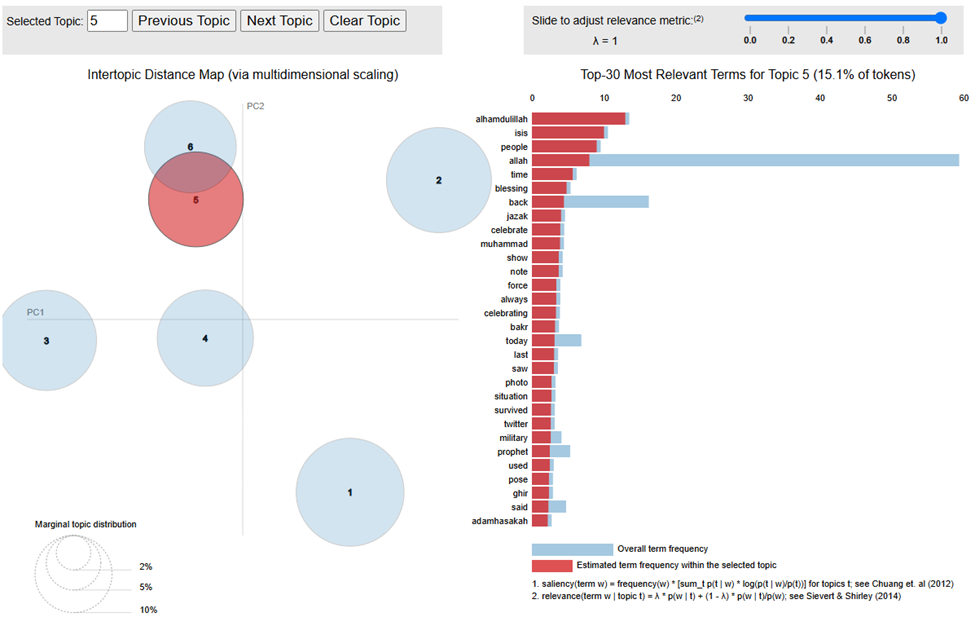
## TOPIC 5

**Community and Pride**

This topic includes **people, Muslim, ummah (community), hope, respect**, and **future**, which are indicative of pride in the **Muslim identity and community**. Words like **ISIS, willayat (province), syrian**, and **amaq** point to specific contexts, but overall, the tone seems to reflect **community pride and optimism for the future**.

**Possible Label**: **Community and Optimism**

*Alhamdulillah, isis, people, allah, time, blessing, back, jazak, celebrate, muhammad, show, note force always, celebrating, bakr, today, last, saw, photo, situation, survived, twitter, military, prophet, used pose ghir, said, adamhasakah*



**Business Use Case and Practical Application:**

In this sentiment-topic tag **positive-community**, we observe many words (people, Muslim, ummah, hope, respect, future) expressing **pride in identity and community solidarity**. These words highlight the effort to **cultivate a sense of belonging and optimism** within the Muslim community, focusing on shared values and collective identity. The term *ummah* specifically promotes the concept of **global brotherhood**, fostering emotional bonds that encourage involvement and loyalty.

Additionally, the inclusion of **contextual references** (ISIS, willayat, Syrian, amaq, Hamas) points to **geopolitical and ideological narratives**. This aligns the sense of community with **regional conflicts and militant struggles**, indicating that ISIS seeks to **link its cause with other extremist movements**. Such alignment with **Hamas**, known for both governance and militant activities, suggests that these messages are not merely intended to nurture sympathizers but also **encourage alignment with violent causes**.

**Positive sentiment indicators** (greatest, nice, celebrating, luck, help) reinforce **emotional engagement**, creating an inviting atmosphere for individuals who may not yet be directly involved but feel drawn toward the ideology. The presence of **visual and multimedia elements** (video, photo) indicates that the narrative leverages engaging content to **spread propaganda and attract a wider audience**.

All in all, this sentiment-topic pair suggests **community-focused messaging** that, while non-violent on the surface, **subtly encourages recruitment into militant networks**. The inclusion of **Hamas** signals an **overlapping ideological framework**, suggesting that ISIS aims to **inspire deeper involvement in regional conflicts** by positioning their cause within the broader struggle for Muslim solidarity.

These tweets create a **sense of shared identity and purpose**, drawing sympathizers toward more active roles. While these messages may not explicitly call for violence, they lay the groundwork for **long-term radicalization**, nurturing supporters who can **amplify propaganda, offer logistical aid, or even participate in future militancy**.

The presence of terms such as *willayat* and *Hamas* underscores the need for **early detection tools** to monitor narratives that connect **identity with militant struggles**. Identifying key influencers and multimedia nodes can reveal **how radical narratives spread through social networks**. Content moderators should track accounts propagating **positive-community messages** that align with extremist goals, as these networks may serve as **recruitment pipelines or logistical hubs for ISIS and affiliated movements**.

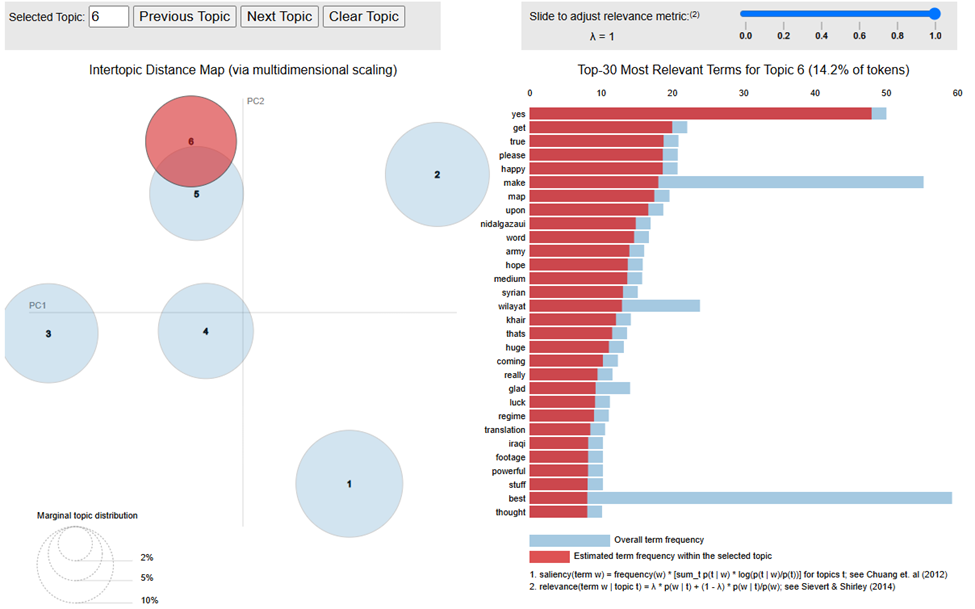
## TOPIC 6

**Enthusiasm and Leadership**

This topic emphasizes words like **great, best, better, thank, epic, protect, and gracious**, which suggest enthusiasm and pride. The terms **Aleppo, Twitter, and Iraq** could point to specific events, but the overarching sentiment seems to involve **leadership, success, and positive outcomes**, both personally and in broader contexts.

**Possible Label**: **Enthusiasm and Success**

*yes get true please happy make map upon nidalgazaui word army hope medium syrian wilayat khair thats huge coming really glad luck regime translation iraqi footage powerful stuff best thought*



**Business Use Case and Practical Application:**

In this sentiment-topic tag **positive-enthusiasm**, the words identified (great, best, better, thank, epic, protect, gracious) suggest **enthusiasm, pride, and motivation**. These terms convey **a positive tone**, celebrating achievements, expressing appreciation, and encouraging further engagement. Words such as *thank* and *gracious* reflect an **atmosphere of recognition** and support, reinforcing emotional connections within the community. The inclusion of **action-oriented terms** (work, note, look, word, field) emphasizes **engagement and participation** within the network, encouraging followers to remain involved in ongoing activities. This messaging paints participation as **valuable and meaningful**, aiming to maintain morale and ensure continuous support. Additionally, **geopolitical references** (Aleppo, Iraq) link the messaging to **ongoing or past conflicts**, portraying these locations as **symbols of struggle and perseverance**. The references evoke emotional responses tied to shared victories, suffering, or resistance in these regions, fostering a sense of **collective pride and purpose**. All in all, this sentiment-topic pair promotes **enthusiasm and positivity** aimed at **sustaining morale** among sympathizers. While not explicitly violent, the messaging creates an environment where involvement and participation are **normalized and encouraged**. The combination of emotional appeal, conflict-related symbols, and interactive content helps to **reinforce solidarity** and maintain the group’s ideological presence.

Content moderators and security agencies can monitor these themes to identify **how positive narratives are used to sustain ideological support**. By tracking content-sharing patterns and identifying key accounts associated with **enthusiastic messaging**, platforms can **disrupt engagement loops** that could lead to further radicalization or logistical support for the group.

# NEUTRAL

## TOPIC 1

**Military Activities (Activities Updates)**

In this sentiment-topic tag **neutral-military activities**, the content revolves around ongoing military operations, specifically in Syria, with terms like **"Syria," "agency," "amaq," "breaking," "army," "Aleppo,"** and **"force."** These terms indicate reports on the military conflict, highlighting battles, attacks, and the involvement of different military groups, such as **"army," "rebel,"** and **"fighter."** There are also references to **"news," "report," "photo,"** and **"video,"** suggesting that these communications serve as battlefield updates and are shared via media channels

**Possible Label**: **Conflict and Military Operations in Syria**

*syria, agency, amaq, breaking, army, aleppo, force, fighter, near, iraq, syrian, iraqi, report, north, rebel, attack, soldier, ramiallolah, today, news, photo, assad, captured, video, ramadi, nidalgazaui, battle, fallujah, position, east*

A screenshot of a graph

Description automatically generated

**Business Use Case and Practical Application:**

**Early Detection of Extremist Messaging**

The analysis of this topic can be instrumental in **early detection of extremist messaging**. The presence of emotionally neutral yet contextually significant terms like **"breaking," "attack," "soldier," "force," "fighter,"** and **"battle"** provides opportunities for **security agencies** to identify when extremist groups are preparing or coordinating military actions. By flagging and analyzing the use of these terms, law enforcement can identify posts that could signal impending attacks, escalating conflicts, or recruitment of fighters. Early detection helps **prevent incidents** by allowing intervention at the right time.

**Improved Content Moderation**

**Social media platforms** such as X, Facebook, and YouTube can leverage this analysis to **improve content moderation strategies**. Instead of simply relying on keyword-based filters, sentiment analysis that includes terms like **"Amaq," "breaking," "news," "captured,"** and **"photo"** can help moderators understand the **context and intent** behind posts. This understanding allows platforms to distinguish between legitimate news reporting and extremist propaganda disguised as conflict updates. Improved content moderation can help reduce the spread of extremist views while ensuring that legitimate information-sharing is not wrongly suppressed.

**Strategic Countermeasures for Law Enforcement and Intelligence Agencies**

**Law enforcement, intelligence, and counter-terrorism agencies** can use sentiment analysis to derive **strategic countermeasures** based on emerging trends. The presence of terms like **"Aleppo," "Ramadi," "rebel,"** and **"position"** allows agencies to understand the **geographical focus** of militant actions and propaganda. By monitoring these terms, they can assess the **shifting interests** of extremist groups and develop pre-emptive measures to counter their plans. These agencies can use such analysis to disrupt recruitment and prevent the spread of influence by understanding the **narrative dynamics** being pushed online.

## TOPIC 2

**Religious and Ideological Support (Support)**

In this sentiment-topic tag **neutral-religious and ideological support**, the content revolves around discussions of **religious themes and ideological support** for certain causes, specifically referencing regions like **Homs, Mosul**, and **Southern** areas. Terms like **"Allah," "Muslim," "Islam," "support,"** and **"Khilafah"** suggest religious sentiments intertwined with the context of the discussions, which could indicate a **religious or ideological justification** for ongoing activities or support for particular actors involved in conflicts. There are references to individuals or user handles such as **"scotsmaninfidel," "sparksOfIrhabi,"** and **"kafirkaty"** that may serve as influencers or key propagators of these messages.

The overall sentiment appears neutral, focusing on ideological discussions, the **promotion of certain beliefs**, and support for actions and events, rather than directly inciting violence or promoting aggressive actions. The content is likely intended to foster **community support**, justify actions through religious context, and provide an **ideological narrative** that ties in regional events, which may serve to engage a broader, ideologically inclined audience.

**Possible Label**: **Religious and Ideological Support**

*allah, homs, muslim, mosul, sparksofirhabi, scotsmaninfidel, time, support, spicylatte, advance, know, back, release, southern, qui, islam, dont, mousaalomar, peigneacheveux, work, khilafah, make, town, sassysassyred, eleven, source, bin, kafirkaty, also, texanna*

A screenshot of a graph

Description automatically generated

**Business Use Case and Practical Application:**

**Early Detection of Extremist Messaging**

The analysis of this topic can provide **early detection of extremist messaging** by highlighting emotionally charged religious and ideological terms like **"Allah," "Islam," "Khilafah,"** and **"support."** These terms, especially when linked to regional conflict areas like **Homs** and **Mosul**, could signal attempts to justify actions using religious ideologies, potentially indicating **early stages of recruitment or radicalization**. By identifying these messages, security agencies can intervene to prevent the spread of extremist narratives.

**Improved Content Moderation**

**Social media platforms** like X, Facebook, and YouTube can utilize this analysis to **improve content moderation** by understanding the religious or ideological context behind posts. Moderation tools that consider terms like **"support," "Islam," "make," "advance,"** and **"release"** can distinguish between legitimate religious discussions and potentially harmful narratives used to promote extremism. This approach would allow platforms to **fine-tune their moderation policies**, reducing the spread of ideological propaganda while avoiding undue censorship of genuine content.

**Strategic Countermeasures for Law Enforcement and Intelligence Agencies**

**Law enforcement and intelligence agencies** can use sentiment analysis to implement **strategic countermeasures** based on the themes detected in these posts. The presence of ideological terms like **"Khilafah," "support,"** and **"Islam"** allows authorities to understand the **narratives being used** to recruit or influence vulnerable individuals. By identifying the **individuals or user handles** such as **"scotsmaninfidel"** and **"kafirkaty"**, agencies can detect **key propagators** and influencers, allowing them to disrupt these networks and mitigate the impact of extremist messaging before it spreads further.

## TOPIC 3

**Military Operations and Governance (Actors)**

In this sentiment-topic tag **neutral-military operations and governance**, the content revolves around discussions of **ongoing military operations** and **government actions** in regions such as **Northern Syria**, **Libya**, and **Palmyra**. Terms like **"Wilayat," "ISIS," "operation," "airstrikes," "government,"** and **"regime"** suggest content related to military and strategic movements involving various regional actors, including **ISIS** and local governments. The presence of terms like **"warreporter," "say," "map,"** and **"line"** indicates that these communications may be updates or discussions on military advancements and territorial control.

The overall sentiment is neutral, focusing on conveying information regarding **military operations, assault lines, and strategic decisions**, rather than promoting an emotional or ideological narrative. The content likely serves as **situational updates**, **battlefield news**, or **reports on government actions** intended for audiences interested in following the developments in conflict zones.

**Possible Label**: **Military Operations and Governance**

*wilayat, isis, northern, new, say, warreporter, libya, regime, palmyra, people, like, deir, may, operation, ezzor, map, help, year, government, line, assault, airstrikes, ahrar, barrack, via, ht, still, twitter, fire, peshmerga*

A screenshot of a graph

Description automatically generated

**Business Use Case and Practical Application:**

**Early Detection of Extremist Messaging**

The analysis of this topic can be utilized for **early detection of extremist messaging** by monitoring terms like **"Wilayat," "ISIS," "operation,"** and **"airstrikes."** These terms indicate military activity or territorial updates involving extremist groups. Identifying such posts helps **security and intelligence agencies** to flag messages that signal **potential recruitment, coordination of military activities**, or spreading extremist narratives, which could be followed by further surveillance and preventive measures.

**Improved Content Moderation**

**Social media platforms** can leverage this analysis to **improve content moderation**, particularly by identifying terms such as **"warreporter," "airstrikes," "assault,"** and **"regime."** These keywords suggest military or extremist content that could be part of ongoing **propaganda efforts**. Content moderators could use this contextual analysis to better distinguish between **legitimate news reports** and **covert attempts at spreading extremist narratives**, thereby enhancing their ability to **reduce harmful content** without impeding genuine information-sharing.

**Strategic Countermeasures for Law Enforcement and Intelligence Agencies**

**Law enforcement, intelligence, and counter-terrorism agencies** can utilize sentiment analysis to derive **strategic countermeasures** by analyzing the discussion of military movements and governance-related actions. The terms **"government," "line," "operation,"** and **"airstrikes"** indicate shifting dynamics on the ground and can help agencies assess which **groups or areas are becoming more active**. By tracking these trends, agencies can plan **proactive interventions** to prevent further escalation, engage in **de-escalation initiatives**, or respond strategically to **emerging threats** in targeted regions.

## TOPIC 4

**Territorial Control and Group Activities (Territorial Control)**

In this sentiment-topic tag **neutral-territorial control and group activities**, the content revolves around discussions related to **territorial control, military activities**, and **group involvement** in specific regions. Terms like **"Islamic," "state," "control," "area," "city,"** and **"village"** suggest discussions about the status of **territory control** by militant groups, particularly the **Islamic State**. Words like **"fight," "group," "join," "fsa,"** and **"YPG"** indicate group activities, alliances, and conflict dynamics in specific areas.

The sentiment seems neutral, focusing on **descriptions of territorial control**, **group movements**, and **battlefield positioning** rather than direct incitement or recruitment. The content is likely intended to provide **situational updates** regarding territorial shifts, alliances, and the status of the areas controlled by different groups, targeting audiences interested in conflict updates and regional power dynamics.

**Possible Label**:

*islamic, state, city, control, area, village, day, fight, group, one, west, medium, countryside, take, apostate, r, see, field, region, last, think, salmanhashimi, fsa, western, join, link, taking, accept, massive, ypg*

A screenshot of a computer

Description automatically generated

**Business Use Case and Practical Application:**

**Early Detection of Extremist Messaging**

The analysis of this topic can assist in **early detection of extremist messaging** by monitoring discussions involving terms like **"Islamic State," "control," "fight," "join,"** and **"YPG."** These terms may indicate efforts to expand territory, recruit fighters, or highlight militant advances. Detecting such language early can allow **security and intelligence agencies** to pre-emptively **monitor and intercept** threats related to territorial gains and the expansion of influence by militant groups, thereby preventing further escalation.

**Improved Content Moderation**

**Social media platforms** such as X, Facebook, and YouTube can use this analysis to **improve content moderation** by identifying terms such as **"control," "fight," "join,"** and **"apostate,"** which suggest extremist activities involving territory and allegiance. Moderation tools can use this contextual understanding to distinguish between **informative content** on regional dynamics and posts that may be veiled attempts to **normalize extremist narratives**. This would help platforms reduce the spread of content that justifies or glorifies militant actions.

**Strategic Countermeasures for Law Enforcement and Intelligence Agencies**

**Law enforcement and counter-terrorism agencies** can use sentiment analysis to plan **strategic countermeasures** based on the evolving status of territorial control and group activities. By monitoring terms like **"village," "control," "fight,"** and **"join,"** agencies can better understand **which regions are under threat** and **how alliances are shifting** between different militant factions. This understanding allows for the formulation of **targeted interventions** in areas at risk and assists in **disrupting alliances** between groups that could strengthen militant movements.

# NEGATIVE

## TOPIC 1

**Conflict and Military Operations**

This topic seems to center around **military conflict**, especially focusing on **army, forces, soldier, attacks, airstrikes, and explosions**. It also mentions specific locations like **Iraq, Fallujah and Aleppo**, which have been key areas of conflict involving various militant groups. The appearance of words like **martyrdom, wounded, explosion and targeted** suggest reports or descriptions of battle outcomes and casualties.

**Negative Label**: **Military Conflict during Attacks**

*killed, Syria, army, iraqi, soldier, amaq, agency, force, near, breaking, destroyed, Syria, city, civilian, assad, child, Iraq, regime, Fallujah, hit, martyrdom, explosion, ramiallolah, tank, Russian, yesterday, wounded, ramadi, bomb, north*

A screenshot of a graph

Description automatically generated

**Business Use Case and Practical Application:**

**Business Use Case:**

Real-time crisis monitoring and response by tracking keywords related to military operations, conflicts, and casualties. Organizations can monitor real-time developments in crisis zones and assist with strategic resource planning. This information is crucial for humanitarian aid organizations, news agencies and governments to coordinate relief efforts to deploy resources and communication tool to the civilian population around the area. Risk assessment for businesses operating in conflict zones or regions with potential for political instability can use this analysis to assess risks to their operations, employees and assets. This information can help them develop contingency plans and make informed decisions about their business strategies.

**Practical Application:**

Humanitarian aid organizations can use this analysis to identify civilian population areas in need of immediate assistance, prioritize resource allocation and coordinate relief efforts. News media outlets can use this analysis to track breaking news stories, verify information to prevent hallucination reporting and provide timely updates to their audience. Government agencies of countries concerning the region can use this analysis to monitor threats, assess potential risks and develop appropriate security measures with allies to cooperate on how to deal with threats.

## TOPIC 2

**ISIS and Civilian Impact**

This topic discusses the relationship between the rebels and civilians **ISIS, Islamic State**, **militant** and **militia**. There’s an emphasis on **kill, claim, attack and clash**. The inclusion of words like **video, today, and airstrike** might imply reports about civilian casualties or the impact of the ongoing conflict on civilians. The terms like **sinai, shiite, palmyra and Aleppo** suggest that this topic could also involve narratives around the areas where the **terrorism** attack occurred and specific ethnic or rebel groups like the Kurds.

**Negative Label**: **Civilian Casualties and Terrorism**

*islamic, state, today, attack, Aleppo, isis, airstrikes, areas, militant, shiite, kill, claim, nidalgazaui, sinai, dead, one, turkey, palmyra, homs, video, fighter, Egyptian, pig, commander, Kurdish, r, Turkish, group, clash, militia*

A screenshot of a computer

Description automatically generated

**Business Use Case and Practical Application:**

**Business Use Case:**

Counter-terrorism intelligence from intelligence agencies can use this topic to track the activities and movements of terrorist organizations like ISIS, identify potential threats, and monitor the spread of extremist ideology. Humanitarian aid and disaster response organizations can use this information to assess the needs of affected populations, plan relief efforts, avoid intense hotspot areas to stay and operate in a safe space and allocate resources effectively.

**Practical Application:**

Intelligence Agencies are able to use this analysis to identify potential targets, disrupt terrorist networks and prevent attacks from occurring. Humanitarian Organization can use this analysis to prioritize aid efforts, identify vulnerable populations and provide targeted assistance. Governments of the countries involved can use this analysis to develop counter-terrorism strategies, strengthen border security and improve intelligence sharing among allies.

## TOPIC 3

**Regime, Rebels, and Fighting**

This topic includes diverse groups of people with similar beliefs with many references to **texanna, kafirkaty, sassysassyred and man**. It seems to highlight key **players in the Syrian conflict** (e.g., **Assad's regime, Israeli/Gaza involvement and Shia fighters**). The topic also mentions **fighting, bombing and death,** indicating the ongoing **war and insurgency**.

**Negative Label**: **Syrian Conflict and Regime Battles**

## *people, scotsmaninfidel, elevn, texanna, bombing, sassysassyred, spicylatte, death, like, allah, kafirkaty, muslim, dont, layer, fighting, leader, know, want, fime, even, make, since, man, never, lost, gaza, get, israel, islam, sparksofirhabi*

A screenshot of a computer

Description automatically generated

**Business Use Case and Practical Application:**

**Business Use Case:**

Geopolitical analysis where concerned governments of countries can use this topic to analyze the complex dynamics of the conflict surrounding the middle east, understand the motivations of various parties to predict potential future developments ahead of these potential threats. Risk assessment the media helps journalists and news organizations use this information to assess the risks associated with reporting from conflict zones and develop safety protocols.

**Practical Application:**

Governments of the countries involved will be able to use this analysis to inform foreign policy decisions, support diplomatic efforts, and provide humanitarian aid. Media companies can use this analysis to develop comprehensive news coverage and identify potential story. Political analysts and researchers can use this analysis to conduct deeper research within this space, publish reports based on findings and influence policy debates towards peace efforts.

## TOPIC 4

**Radical Ideology and Western Opposition**

This topic features references to **rebel and terrorist**. It includes **Western figures** or mentions like **police, warreporter, and kafir** (a derogatory term for non-believers in radical contexts). The terms **lqaeda, ahrar, apostate, kuffar, adamhasak** may indicate radical propaganda or extremist content aimed at rallying support against Western nations and ideologies.

**Negative Label**: **Extremism and Western Opposition**

*rebel, village, killing, warreporter, police, terrorist, apostate, ypg, coalition, body, muslim, saudi, take, air, strike, adamhasak, ahrar, vso, http, another, arab, medium, alqaeda, kuffar, kurd, fsa, dog, crime, afp*

A screenshot of a graph

Description automatically generated

**Business Use Case and Practical Application:**

**Business Use Case:**

Counter-terrorism intelligence agencies can use this topic to identify extremist groups and individuals, track their activities, and disrupt their networks. Cybersecurity firms can use this information to analyze by identifying potential cyber threats from extremist groups and develop strategies to protect critical infrastructure in a timely intervention order.

**Practical Application:**

Intelligence agencies are able to use this analysis to identify potential threats, disrupt terrorist networks to potentially prevent attacks. Cybersecurity firms can use this analysis to develop threat intelligence, improve security measures, protect critical infrastructure with precautionary measures. Social Media Platforms can use this analysis to moderate, identify and remove extremist content, disrupt recruitment efforts and prevent the spread of hate speech.

## Appendix 5

**Semi-Supervised LDA output**

# POSITIVE

## TOPIC 1

**Faith and Resilience in Challenging Times**

This topic represents tweets that combine elements of faith, resilience, and hope during challenging times, often using religious expressions and references to places affected by conflict. Words like ***allah, ameen, accept, worship, and true*** highlight a religious or spiritual context. Terms such as ***syria, mosul, operation, state, and better*** suggest a focus on current events, especially in regions experiencing hardships. There’s an emphasis on improvement and resilience with words like victory, better, strong, and always, which may imply a sentiment of enduring hope for positive change amid difficulties.

**Positive Label**: **Hopeful Resilience**

*allah, may, beautiful, day, see, islamic, state, make, victory, pleased, get, funny, ameen, true, accept, syria, better, upon, look, come, well, mosul, medium, picture, go, yes, worship, always, strong, operation*

**A screenshot of a graph

Description automatically generated**

**Business Use Case:**

Topic 1 is relevant for businesses and organizations that aim to engage with audiences on values of community support and resilience. Social media platforms can monitor this sentiment to guide content moderation and crisis response strategies, while nonprofits can design culturally resonant campaigns that connect with affected communities. Brands, especially in health and wellness, can craft empathetic messaging that aligns with these values, deepening connections with audiences who prioritize resilience and faith. Corporate Social Responsibility (CSR) initiatives can incorporate these themes to foster trust and make meaningful contributions in regions experiencing hardship. Media outlets, too, can leverage this focus to create positive narratives that promote global solidarity and encourage community strength in challenging times.

## TOPIC 2

**Social Media Engagement and Global Events**

This topic reflects conversations around current global events, particularly those related to conflict zones, social media updates, and engagement on Twitter. The terms include mentions of regions like ***Iraq and Syria***, as well as references to ***ISIS and military forces***, indicating discussions of geopolitical situations. There is also a focus on media elements like ***video, photo, and map***, which suggests that users are sharing content or updates about these events. Terms like ***good, thank, and happy*** imply that, despite serious subject matter, these conversations may include positive engagement or expressions of solidarity.

**Positive Label**: **Global Awareness and Solidarity**

**A screenshot of a graph

Description automatically generated***Good, best, isis, video, thank, week, Iraq, came, great, rts, happy, back, follow, also, map, make, like, nidalgazaui, word, note, Syrian, force, soon, huge, significant, coming, photo, situation, military, twitter*

**Business Use Case:**

This topic is especially useful to organizations tracking global events and public sentiment, such as news agencies and social media platforms. They can enable targeted content distribution and effective misinformation countering. Social media companies can also use these insights for contextual advertising, reaching audiences interested in global affairs with relevant ads like news subscriptions and humanitarian aid. Government and security agencies can analyze this data to assess security risks and public perception in conflict regions, supporting public safety efforts.

## TOPIC 3

**Expressions of Faith, Support, and Community Connection**

This topic revolves around expressions of **faith, community support, and positive social interactions** within a religious or spiritual context. The use of terms **like Allah, Muhammad, ummah, akhi, and alhamdulillah** emphasizes a faith-based language, likely associated with Muslim communities. Terms such as **love, amazing, support, and care** indicate positive emotions and support within the community, while **fallujah, khilafah, and wilayat** might reference broader cultural or geopolitical themes. The presence of words like **thank, fun, and nice** adds a lighter tone, suggesting casual, friendly exchanges or expressions of gratitude.

**Positive Label**: Community Unity and Spiritual Support

*wilayat, isis, northern, new, say, warreporter, libya, regime, palmyra, people, like, deir, may, operation, ezzor, map, help, year, government, line, assault, airstrikes, ahrar, barrack, via, ht, still, twitter, fire, peshmerga*

*A screenshot of a graph

Description automatically generated*

**Business Use Case:**

Social media platforms can enhance user experience by promoting positive, supportive content that fosters community belonging, while brands can leverage this understanding to create culturally sensitive campaigns that emphasize values like family, unity, and positivity. Nonprofit organizations can use these insights to tailor communication strategies that align with community values, improving outreach for welfare and humanitarian causes. Government agencies and NGOs can incorporate these insights to design crisis response and community support initiatives, strengthening resilience in faith-based regions. Faith-focused media outlets can also benefit by crafting content that resonates with themes of support and positivity, inspiring their audience and enhancing engagement with faith-centered narratives.

## TOPIC 4

**Expressions of Positivity, Resilience, and Support**

This topic centers around expressions of **positive sentiments, resilience, and support**, particularly in the context of global or community issues. Words like **good, thanks, welcome**, and best suggest a positive tone, while terms such as **support, freedom, resilience, and better** imply a spirit of perseverance or encouragement. The mentions of specific places and terms like **Syria, Aleppo, and Quran** indicate that these messages might resonate with themes of cultural or regional support, resilience in adversity, and religious or moral encouragement. This topic reflects a blend of personal support, acknowledgment of hardships, and a positive outlook.

**Positive Label**: Community Unity and Spiritual Support

*good, one, thanks, welcome, Syria, better, think, indeed, cool, keep, best, brother, ramiallolah, know, aleppo, akhi, world, give, much, guy, power, yeah, freedom, support, night, interesting, dying, still, Quran, year*

**A screenshot of a graph

Description automatically generated**

**Business Use Case:**

Social media platforms can use these insights to promote uplifting content and recommend groups that focus on resilience and mutual support, especially in regions facing challenges. Brands in health, wellness, and social causes can craft messages around community strength and mental resilience, creating a stronger emotional bond with audiences. Nonprofits and CSR initiatives can leverage these themes to develop culturally sensitive campaigns that resonate with communities in hardship, inspiring hope and solidarity. Media outlets can use these insights to produce stories of resilience and unity, while community organizations can organize events that emphasize support, cultural pride, and collective resilience, enhancing a sense of belonging within faith-based or cultural groups.

## TOPIC 5

**Expressions of Faith, Respect, Understanding**

This topic focuses on expressions of faith, respect, and understanding, often conveyed through religious phrases and positive sentiments. The use of terms like **Allah, alhamdulillah, subhanallah**, **and bless** suggests a theme of reverence and gratitude in a religious context. Words like **people, respect, protect, and understand** imply a focus on empathy and mutual respect, while terms like **great, love, and amazing** reflect positivity and admiration. There is also an emphasis on learning and comprehension with words such as **understand, learn, and translation**, suggesting an interest in sharing or interpreting values and teachings.

**Positive Label**: Reverence and Mutual Respect

*Allah, people, great, yes, time, love, alhamdulillah (praise be to God), please, may, sparksofirhabi, celebrate, sheikh, bless, work, subhanallah (glory be to God), khair (good), protect, understand, lean, bakr, reward, respect, amazing, right, translation, thought.*

A screenshot of a graph

Description automatically generated

**Business Use Case:**

The topic highlights sentiments of reverence, gratitude, empathy, and a desire for mutual respect. Social media platforms can use these insights to engage faith-based communities by promoting uplifting and respectful content that fosters cultural sensitivity. Brands can leverage this understanding to connect with culturally diverse audiences through campaigns that align values of respect, gratitude, and community. Additionally, public relations teams can use this analysis to tailor culturally sensitive responses and build trust with diverse audiences, especially during sensitive or challenging times.

# NEUTRAL

## TOPIC 1

**Support and Group Dynamics (Support)**

In this sentiment-topic tag **neutral-support and group dynamics**, the content revolves around **support for ISIS**, discussions of group activities, and mentions of key actors involved. Terms like **"ISIS," "Ramiallolah," "Muslim," "fight," "support,"** and **"group"** suggest that the content is centered around **group dynamics, solidarity, and support** related to ISIS activities and operations. There are references to individuals such as **"Ramiallolah"** and user handles like **"sparksofIrhabi"**, which could indicate key propagators or influential figures involved in these discussions.

The sentiment appears neutral, focusing on **group actions, support mechanisms**, and **operational updates** without expressing overt emotional content. The content likely serves as **group cohesion reinforcement**, updates on activities, or statements of support for ISIS and its affiliates, targeting an audience that is sympathetic or interested in following these developments.

**Possible Label**: **Conflict and Military Operations in Syria**

*isis, ramiallolah, new, say, muslim, warreporter, fight, one, sparksofIrhabi, people, like, time, group, support, assad, saudi, know, day, see, help, islam, dont, year, government, mousaalomar, ahrar, via, israel, twitter, make*

A screenshot of a computer

Description automatically generated

**Business Use Case:**

**Early Detection of Extremist Messaging**

The analysis of this topic is useful for **early detection of extremist messaging** by identifying discussions involving terms like **"ISIS," "Ramiallolah," "support,"** and **"fight."** These terms indicate **group solidarity** and potential recruitment or mobilization efforts, which could escalate into coordinated actions. By flagging such terms early, **security agencies** can identify emerging threats and take preemptive action to counter **recruitment** or **coordination** among ISIS supporters.

**Improved Content Moderation**

**Social media platforms** such as X, Facebook, and YouTube can leverage this analysis to **improve content moderation** by identifying and understanding the context behind terms like **"group," "support," "Islam,"** and **"fight."** Such analysis can help platforms distinguish between genuine religious discussions and **extremist group dynamics** meant to rally support. By incorporating context into their moderation policies, platforms can better address **harmful narratives** while reducing the risk of **over-censorship** of legitimate content.

**Strategic Countermeasures for Law Enforcement and Intelligence Agencies**

**Law enforcement and intelligence agencies** can use sentiment analysis to implement **strategic countermeasures** by focusing on the **group dynamics** and **support systems** detected in this content. Terms like **"support," "group," "fight,"** and **"Ramiallolah"** suggest strong group cohesion and internal reinforcement of beliefs. Agencies can use this insight to plan **targeted interventions**, such as disrupting **key propagators** and **influential figures** (e.g., "sparksofIrhabi") that are instrumental in maintaining group cohesion and influencing others, ultimately preventing the further spread of extremist narratives.

## TOPIC 2

**Military and Territorial Updates (Updates)**

In this sentiment-topic tag **neutral-military and territorial updates**, the content focuses on **military activities**, **territorial control**, and **ongoing operations** in conflict areas like **Syria, Aleppo**, and **Iraq**. Terms like **"Islamic State," "Wilayat," "army," "force," "fighter," "attack,"** and **"control"** indicate military operations and the efforts of different groups, including ISIS, to maintain or gain territorial control. References to **"breaking," "report," "news," "captured,"** and **"video"** suggest that these are updates on ongoing conflicts and battles, providing insights into the status of areas and military advancements.

The overall sentiment appears neutral, providing **situational reports**, **battle updates**, and **territorial information** without strong emotional connotations. The content is likely intended for audiences interested in **following the progress of military activities**, **territorial disputes**, or **group movements** within the affected regions.

**Possible Label**: **Religious and Ideological Support**

*islamic, state, syria, wilayat, agency, amaq, breaking, army, aleppo, force, city, fighter, near, iraq, iraqi, report, north, homs, attack, soldier, northern, control, syrian, news, photo, area, captured, rebel, village, video*

A screenshot of a graph

Description automatically generated

**Business Use Case:**

**Early Detection of Extremist Messaging**

The analysis of this topic can be used for **early detection of extremist messaging** by monitoring key terms like **"Islamic State," "Wilayat," "attack,"** and **"force."** These terms suggest that **territorial and military advancements** are being reported or discussed by extremist groups. By identifying such terms in real time, **security agencies** can gain insights into **planned or ongoing attacks** and intervene to prevent further escalation or recruitment of fighters, particularly in areas like **Syria** and **Iraq**.

**Improved Content Moderation**

**Social media platforms** can use this analysis to **improve content moderation** by recognizing terms like **"Wilayat," "army," "rebel," "force,"** and **"video."** Such words indicate the presence of content related to **military actions and territorial control**, potentially serving as propaganda. Moderators can better understand the **context** and identify posts that may be disguised as news but actually serve to **promote extremist narratives**. This will help ensure that such content is appropriately flagged and removed, while legitimate news content is not overly censored.

**Strategic Countermeasures for Law Enforcement and Intelligence Agencies**

**Law enforcement and intelligence agencies** can use sentiment analysis to plan **strategic countermeasures** by analyzing the geographic and military-related terms, such as **"control," "captured," "Aleppo," "Homs," "report,"** and **"news."** These terms help agencies understand **shifts in territorial control** and which regions are becoming active conflict zones. By monitoring these terms, intelligence agencies can implement **targeted interventions** in regions at risk, predict **future areas of conflict**, and plan **preventive measures** to disrupt the influence of extremist groups like ISIS.

## TOPIC 3

**Religious and Social (Religious)**

In this sentiment-topic tag **neutral-religious and social commentary**, the content revolves around discussions that include **religious references**, **user interactions**, and **social dynamics**. Terms like **"Allah," "brother," "ameen,"** and **"akhi"** suggest a focus on **religious themes**, while mentions of **"prison," "link," "protect,"** and **"watch"** imply discussions around social or political events, possibly related to the well-being of specific individuals or groups. Additionally, user handles such as **"scotsmaninfidel," "spicylatte,"** and **"sassyassyred"** indicate engagement from identifiable figures or influencers, contributing to these discussions.

The sentiment seems neutral, focusing on **religious expressions**, **social observations**, and **commentary on political or personal matters** without direct emotional incitement. The content is likely intended for audiences interested in **social and religious commentary**, with discussions that revolve around individual experiences, religious solidarity, and reactions to specific incidents.

**Possible Label**: **Military Operations and Governance**

*allah, scotsmaninfidel, spicylatte, eleven, sassyassyred, texanna, may, kafirkaty, back, peigneacheveux, follow, brother, l, prison, link, accept, please, pour, fukuyamazigh, h, wa, aussi, akhi, non, ibn, ameen, mais, protect, watch, fasabrunjameel*

A screenshot of a graph

Description automatically generated

**Business Use Case:**

**Early Detection of Extremist Messaging**

The analysis of this topic is useful for **early detection of extremist messaging** by monitoring the **religious language** and user interactions, such as terms like **"Allah," "brother," "akhi,"** and **"protect."** These terms, combined with mentions of **"prison"** or **"link,"** could imply support or solidarity with individuals linked to extremist activities or discussions about incarceration of group members. **Security agencies** can use this information to identify potential **calls for action** or **messages of support** for imprisoned radicals, which could indicate early stages of mobilization or radicalization.

**Improved Content Moderation**

**Social media platforms** can leverage this analysis to **improve content moderation** by identifying terms such as **"protect," "watch," "prison,"** and **"accept,"** which may indicate **social or political discussions** involving sensitive topics. Moderators can analyze the context behind these discussions to determine if the content is being used to subtly **promote extremist narratives** or **incite action**. By distinguishing between **genuine religious expressions** and **hidden extremist undertones**, content moderation can be more nuanced, reducing the spread of harmful content while protecting legitimate discussion.

**Strategic Countermeasures for Law Enforcement and Intelligence Agencies**

**Law enforcement and intelligence agencies** can use sentiment analysis to derive **strategic countermeasures** by monitoring user handles and terms that imply **influencer activity** and **social support** within certain communities. Identifying individuals like **"scotsmaninfidel"** and **"sassyassyred"** and tracking the use of phrases such as **"brother," "protect,"** and **"watch"** can help understand **influence dynamics** within these groups. These insights can be used to plan **targeted interventions** against key propagators or influencers to **disrupt networks** and mitigate the spread of ideologies that could lead to radicalization.

# NEGATIVE

# SEMI-SUPERVISED LDA NEGATIVE

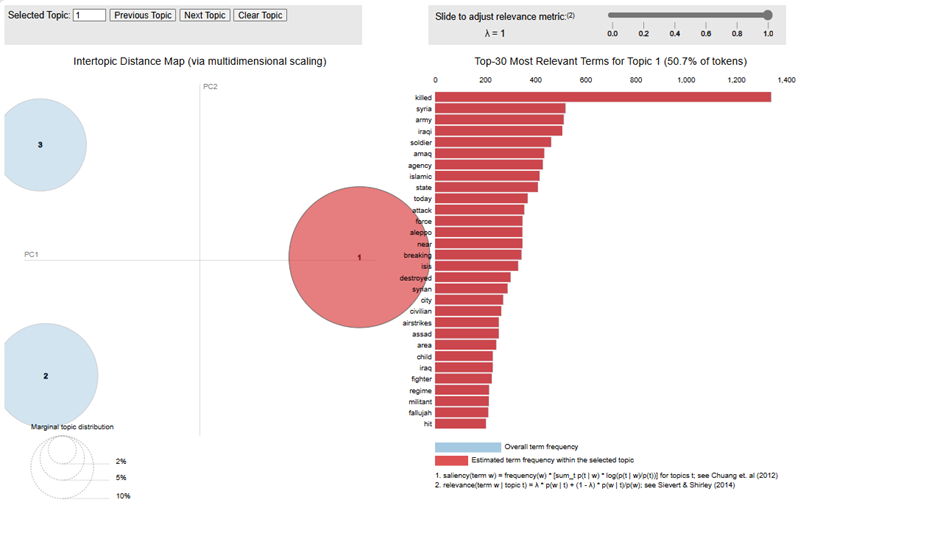
## TOPIC 1

**Military Conflict during Attacks**

This topic centers on military conflict focuses on words like **"killed," "fighter," "army," "soldier," and "airstrikes,"** suggesting ongoing warfare and casualties in specific regions like Iraq, Syria, and Fallujah. The emphasis of words like **"martyrdom," "explosion," “destroyed” and "wounded"** indicate severe battle outcomes and damage in conflict zones. This language reflects reports on military operations and the aftermath, likely providing updates following the outcomes of these conflicts.

**Negative Label**: **Warfare and Combat Casualties**

*killed, syria, army, iraqi, soldier, amaq, agency, islamic, state, today, attack, force, aleppo, near, breaking, isis, destroyed, syrian, city, civilian, airstrikes, assad, area, child, iraq, fighter, regime, militant, fallujah, hit*



**Business Use Case:**

This analysis can be applied to real-time crisis monitoring and response, as organizations can track keywords related to military actions and assess the impact on civilians and infrastructure. For example, humanitarian agencies can identify areas requiring urgent aid, while news organizations might use this analysis to provide timely and verified updates on the ground situation. Furthermore, government agencies can leverage these insights to assess threats in real-time, aiding in the coordination of security measures and deployment of resources.

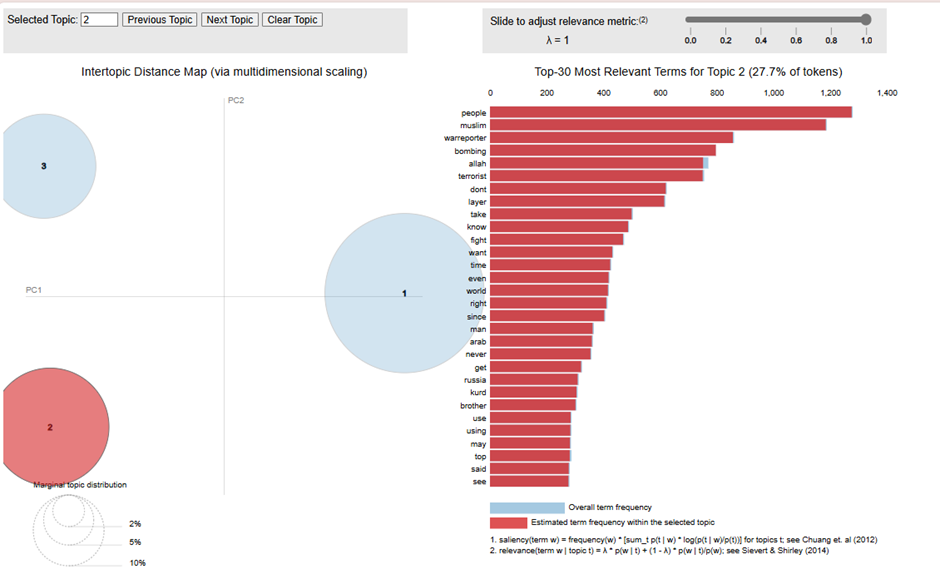
## TOPIC 2

**Civilian Casualties from Terrorism Activities**

This topic reflects the impact of ISIS activities on civilians, with key terms including **"people," "muslim," "warreporter," "bombing," and "terrorist."** The presence of words like **"fight," "want," "world," and "kurd"** suggests themes related to religious and ideological tensions. High-risk locations like Aleppo and Sinai likely feature prominently in this narrative, with terms like **"take," "man," and "brother"** indicating the toll of militant actions on individuals and communities.

**Negative Label**: **Terrorism Impact on Civilians**

*people, muslim, warreporter, bombing, allah, terrorist, dont, layer, take, know, fight, want, time, even, world, right, since, man, arab, never, get, russia, kurd, brother, use, using, may, top, said, see*



**Business Use Case:**

Counter-terrorism intelligence can benefit significantly from tracking these discussions to detect potential threats early. For instance, intelligence agencies can use this analysis to monitor extremist activities, evaluate the spread of militant influence and prevent recruitment efforts. Humanitarian organizations could use these insights to prioritize aid and ensure the safety of affected civilians in high-risk areas. Similarly, government agencies may integrate this analysis to develop robust counter-terrorism strategies, bolster border security and promote intelligence sharing across high-risk regions vulnerable to extremist attacks.

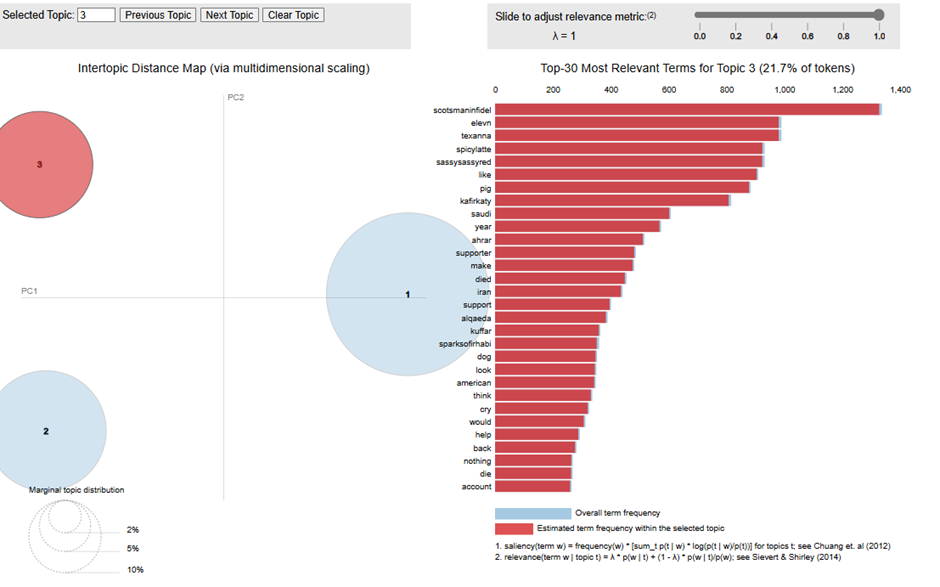
## TOPIC 3

**Ideological Influenced Regime Battles**

This topic revolves around the Syrian conflict, highlighting figures and groups such as Assad's regime, Shia fighters and other rebel forces like Al-Qaeda. Key terms include **"scotsmaninfidel," "texanna," "ahrar," “kuffar,” and "kafirkaty,"** which suggest discussions influenced by religious or ideological divides. Words like **"fighting," "bombing," "death," “cry,” and "died"** emphasize the conflict's violent nature and setbacks that occurred in the aftermath of battle, while terms like **"support," "back," and "help"** indicate both support and resistance within the conflict's social framework and sympathizers who back the rebel forces.

**Negative Label**: **Ideological Conflicts in Syria**

*scotsmaninfidel, elevn, texanna, spicylatte, sassysassyred, like, pig, kafirkaty, saudi, year, ahrar, supporter, make, died, iran, support, alqaeda, kuffar, sparksofirhabi, dog, look, american, think, cry, would, help, back, nothing, die, account*



**Business Use Case:**

Geopolitical analysis agencies can leverage this information to track the Syrian conflict's evolving dynamics, understand various actors' motivations and predict potential developments. Media organizations could use these insights to assess reporting risks and develop safety measures for journalists in the field. Political analysts and research institutions may find this analysis valuable for producing comprehensive reports on conflict drivers, informing foreign policy to efficiently allocate resources and guiding humanitarian efforts. Geopolitical analysis agencies can leverage this information to track the Syrian conflict's evolving dynamics, understand various parties’ motivations and predict potential developments.

## Appendix 6

## Group reflection

A lack of expertise in selecting seed words for semi-supervised LDA can present challenges. Since most team members may not be familiar with the topics of ISIS and terrorism, certain keywords with significant contextual meanings may not be captured effectively. This can lead to an incomplete representation of the intended topics, as key terms are missed during seed word identification and allocation.

Deciding how to handle foreign language content is also crucial. While removing foreign language text from the corpus may seem simplest, it risks excluding words with significant meaning. Moreover, using general machine translation for specific topics like terrorism may not yield ideal results, as nuances in terminology may be lost, reducing the accuracy of topic analysis.

## Individual Reflections

|  |  |
| --- | --- |
| Member | Reflection |
| Goh Cheng En | I was able to better understand the different types of models for sentiment analysis and appreciate how it could be applied at workplaces. I contributed by coming with the TF-IDF seed words and collaborated on the seed topics for the business and practical use cases that was used for the Semi-supervised LDA model. Then I did the write up for the business used cases and practical applications on neutral sentiments words for both semi-supervise LDA and unsupervised LDA models. I eventually stitched up everyone’s part in the report and through the stitching was able to better appreciate the discussions and individual work that was going on. It really highlighted the importance of learning from one another even as we split up the tasks so that it does not become overly silo and missed out on the overall flow of the project. Through the use of this dataset, I have also appreciated the fact that often times there might not be a ground truth to sentiment analysis in the real world context and may need to create one manually or sentiment models with confidence scores. |
| Low Wei Jun | This project gave me a valuable insight as to how leveraging text analytics can be applied into a sensitive subject matter which is relevant to the real world scenario. Applying what I learnt in text analytics, I contributed to description in the sentiment analysis, potential use cases and the negative sentiments in the LDA semi-supervised and unsupervised topic modeling process, assisted in the summarization of both drafts and final reports and the word clouds for the semi-supervised LDA topic modelling and from identifying business use cases, generating TF-IDF seed words for negative sentiment analysis, and writing on the practical applications of our findings and thereafter prepared the presentation word clouds for the semi-supervised LDA topic modelling. Working with a diverse team from various work experience backgrounds and perspectives, we adopted the agile methodology approach to collaborating effectively to overcome challenges, such as deciding on supervised LDA modeling and harmonizing our collective efforts to be summarized into the report and the presentation slides. Ultimately, we each utilized our knowledge to breakdown the work on each task to each of our own strengths and managed the project deliverables efficiently and promptly. This project was truly an educational experience for me in understanding the importance of data analytics and on the business side, crafting topics relevant to real-life scenarios and handling text analytics on sensitive subjects. Despite initial hesitations about analyzing tweets on complex subjects with limited context, our curiosity and determination to tackle this challenging topic led us to valuable insights. |
| Lorielle Malveda | Through this project, I gained valuable insights into the power of text analytics, particularly in analyzing the ISIS tweets dataset to uncover meaningful patterns. My contributions included interpreting the metrics to evaluate the topic modeling, determining the optimal number of topics, interpreting the words generated by the model, and coming up with relevant topics, which enhanced my understanding of both supervised and unsupervised topic modeling approaches. I enjoyed applying text analytics in a real-world context, although interpreting cultural nuances and specific jargon in the dataset proved challenging and required extra research. Overall, this project was both challenging and enriching, highlighting the importance of domain knowledge in data analysis. |
| Sarthak Nagapurkar | This project provided me with valuable learning experience in hands-on implementation of unsupervised and semi-supervised topic modeling. Initially, there were challenges in running topic modeling on my machine due to gensim version incompatibility but after having navigated those challenges I realized how important data quality was when I identified noise was biasing the topic modeling results. During the implementation of semi-supervised LDA, I realized that the identification of appropriate seed topics and seed words aligned with the project objectives was crucial. The other team members with experience and understanding of business use cases pitched in to help in that part. |
| Huang Shaofei | This project offered a valuable opportunity to deepen my skills in sentiment analysis, topic modeling, and collaborative work. I particularly enjoyed the hands-on aspect of developing a sentiment analysis pipeline, which provided useful insights into the strengths of rule-based versus deep learning approaches in text analysis. Collaborating with the team was highly rewarding, as each member’s unique skills strengthened our approach, although aligning schedules and balancing different work styles could be challenging at times. Overall, while the project enriched my technical knowledge and collaborative abilities, I found the occasional misalignment in pacing and expectations to be a drawback. |
| Liang Xiuhao | I think that we are a group with different expertises, some in coding, research and subject matter like terrorism. The group is receptive to try new ideas when we hit roadblock in the supervised LDA, but nonetheless, we tried different methods and collectively solve the issue. |