SENG 550 – Project

2024-12-19

Sentiment Analysis on Queen Elizabeth II’s Death

Aarushi Roy Choudhury: 30113987

Farica Mago: 30111924

Shanelle Li Chit Khim: 30124614

Contents

Preamble

Team Member Contributions

- Aarushi:
  - Worked on building the unsupervised models with K-Means clustering, LDA topic modeling, lexicon and rule-based models with tools like VADER and TextBlob.
  - Built the data cleaning and transformation pipelines to extract features for the ML models.
  - Contributed to the report and video demo.
  - Estimated contribution: 33%
- Farica:
  - Worked on logistic regression model with rule-based heuristics sentiment labeling approach
  - Contributed to the report and video demo.
  - Estimated contribution: 33%
- Shanelle:
  - Worked on evaluating unsupervised models with metrics such as UMass and evaluating random sample predictions through human inspection.
  - Contributed to the report and video demo.
  - Estimated contribution: 33%

Declaration signed by all members that the statement of contributions and estimate of total contribution is true:

Aarushi, Farica, Shanelle

Abstract

This project analyzes public sentiment and discussions about the passing of Queen Elizabeth II through a dataset of tweets. By employing unsupervised learning methods like K-Means clustering and Latent Dirichlet Allocation (LDA), sentiment analysis tools (VADER, TextBlob), and supervised learning with logistic regression, we explored global emotional responses and societal reflections on this historic event. Key findings include the identification of prevalent topics, sentiment trends, and the effectiveness of heuristic-based supervised learning in the absence of labeled data. Challenges in dataset quality and cluster coherence were addressed, providing insights into the complexities of analyzing social media data.

Introduction

What is the Problem?

The problem is analyzing and understanding public sentiment and discussions about the death of Queen Elizabeth II. The dataset consists of tweets from around the world, capturing various opinions, emotional reactions, and perspectives in the wake of her passing. By conducting sentiment analysis and statistical modeling, our goal is to uncover trends in how people expressed their feelings, identify the key themes and topics discussed, and potentially predict the type of sentiment expressed based on specific tweet characteristics. This analysis aims to provide insights into global emotional responses and societal views during a significant historical event.

Why is it an Important Problem?

The death of Queen Elizabeth II marked a major historical event with profound emotional, cultural, and political implications worldwide. Analyzing public sentiment through social media, particularly Twitter, provides a unique opportunity to understand how people from different backgrounds, regions, and political beliefs reacted to her passing. The royal family, including Queen Elizabeth II, has had a significant influence on public opinion due to their involvement in key historical events such as colonization, as well as their continued role in modern society. While some expressed grief and respect for her leadership, others voiced criticisms related to the royal family's legacy of colonialism and its lasting impacts.

By examining these sentiments, we can uncover trends in emotional responses, identify key themes discussed, and assess the broader societal views shaped by this complex history. Such analysis helps in understanding how major events like the death of a prominent figure influence global public opinion, highlight divisions, and potentially reveal calls for accountability regarding past actions. This type of analysis is crucial for governments, media, and organizations to better respond to public concerns and emotions, while also contributing to our understanding of how historical narratives influence contemporary society.

What have Others Done in this Space?

Previous studies have analyzed social media sentiment in contexts such as political events, pandemics, and high-profile deaths. For example, Chintalapudi et al. (2021) analyzed COVID-19 tweets to identify emotional patterns. However, sentiment analysis on Queen Elizabeth II's passing remains limited, with most work focusing on descriptive clustering without exploring nuanced themes like mourning or geopolitical critique.

What are Some Existing Gaps that we Seek to Fill?

While studies, such as the one in Journal of Language, Literature, Social, and Cultural Studies (2024), have conducted content and sentiment analyses on public reactions to Queen Elizabeth II's death on Twitter, several gaps remain that our work aims to address:

1. **Integration of Supervised and Unsupervised Learning:**
   - Existing Research: Prior analyses predominantly utilize either sentiment analysis tools or clustering methods in isolation.
   - Our Contribution: We combine unsupervised learning techniques, such as K-Means clustering and LDA topic modeling, with supervised learning approaches, including logistic regression, to enhance the robustness and depth of our analysis.
2. **Heuristic-Based Labeling for Supervised Models:**
   - Existing Research: The absence of labeled data has limited the application of supervised learning models in this context.
   - Our Contribution: We develop heuristic rules to generate labels for training a logistic regression model, enabling predictive analysis of tweet sentiments in the absence of pre-labeled datasets.
3. **Comparative Analysis of Sentiment Analysis Tools:**
   - Existing Research: Limited studies compare the effectiveness of different sentiment analysis tools on this specific dataset.
   - Our Contribution: We conduct a comparative evaluation of VADER and TextBlob, assessing their performance and highlighting the nuances in their sentiment classification, particularly in detecting mourning sentiments.

What are our Data Analysis Questions?

- What are the predominant sentiments (positive, negative, neutral) expressed in tweets about Queen Elizabeth II's death?
- Can machine learning models predict the sentiment of tweets based on their content, and how accurate are these models?
- What are the key themes and topics discussed in the tweets, and how do they reflect public opinion on the royal family's legacy?
- Are there discernible patterns in the way individuals discuss historical events related to the monarchy (e.g., colonization) versus personal or emotional reflections on her death?

What are we Proposing and what are our Main Findings?

We proposed using unsupervised learning with K-Means clustering, topic modeling with LDA; sentiment analysis tools (VADER, TextBlob); and supervised learning with logistic regression and heuristic labels. We compared the results supervised vs. unsupervised learning. Key findings include:

- K-Means clustering and LDA revealed incoherent clusters and topics due to poor data quality.
- A major limitation of our analysis was the dataset quality, which contained minimal sentimental data and repetitive phrases, such as variations of "Queen Elizabeth," "II”. These dominant words appeared across clusters, making it challenging to extract meaningful insights or perform distinct topic modeling.
- Although sentiment analysis tools like VADER and TextBlob were applied, their effectiveness was limited by the lack of granular, labeled data. Furthermore, VADER and TextBlob do not understand the concept of “mourning” which can change the context positive vs. negative sentiments.
- VADER and TextBlob performed well with heuristics but faced domain-specific challenges.
- Logistic regression achieved high accuracy with supervised learning on simple heuristic-labeled data.

Methodology

Refinement and Exploration of Data Features

- **Data Cleaning:**
  - Removed URLs, mentions, and stopwords.
  - Tokenized and lemmatized text to normalize word forms.
  - Engineered features using CountVectorizer for clustering and LDA.
- **Exploratory Data Analysis:**
  - Identified the top 10 most frequent words to get a sense of the data.

Experimentation Factors

1. **Unsupervised Learning:**
    - K-Means clustering to identify thematic groupings.
        - silhouette score used to tune hyperparameter ‘k’ and find optimal ‘k’ value
    - LDA for topic modeling, evaluating coherence scores.
2. **Lexicon and Rule-Based Tools**:
    - VADER for nuanced social media sentiment.
    - TextBlob for polarity-based sentiment.
    - Custom "mourning" detection added.
3. **Supervised Learning:**
    - Logistic regression trained on heuristic-labeled data (positive, negative, neutral, mourning).
    - Data split into training and test set.

Experiment Process

The dataset has over 190K data points with 36 columns, so we used Spark to load, read, process and analyze the data. We used Google Colab to access Spark. We installed Spark, set up the environment variables, and initialized a SparkSession in the Colab notebook environment. This approach allowed us to combine the scalability and distribution capabilities of Spark with the convenience, interactivity, and rich Python ecosystem available in Colab.

One of the critical challenges we faced is the absence of "true labels" in our dataset. The only way to generate true labels would be to manually label a sample of 100-500 tweets with sentiments based on human judgement. This is time consuming, expensive and out of scope for the project. To address this, we explored methods based on unsupervised learning techniques, and we leveraged sentiment analysis tools like VADER and TextBlob.

Unsupervised Learning

1. **Data Cleaning**
    - We cleaned the dataset by removing empty tweets, non-English tweets, duplicate tweets, upper case text, URLs, mentions (@handles), hashtags (#), and punctuation.
2. **Feature Engineering**
    - We converted raw tweets into a processed format using Spark NLP annotators. This includes tokenization, lemmatization (to normalize words like "running," "runs," and "ran" to their root "run"), and removal of noise to ensure higher cluster coherence. The processed tokens are then converted into numerical features using CountVectorizer, enabling machine learning algorithms to handle the data.

3. **Exploratory Data Analysis**
    - To understand the dataset better, we identified the top 10 most frequent words in the processed tweets. This provides insights into key themes and the overall focus of the conversations.
4. **K-Means Clustering**
    - As the dataset is not labelled i.e. for each tweet there is no pre-defined sentiment label, we tried K-Means clustering first to discover thematic groups (mourning, political critiques, historical context).
    - We analyzed the clusters found and inspected to see if any patterns or themes appeared.
5. **Topic Modeling with LDA**
    - We also explored the topics found by LDA and compared it to K-Means.
    - In natural language processing, latent Dirichlet allocation (LDA) is a Bayesian network (and, therefore, a generative statistical model) for modeling automatically extracted topics in textual corpora. The LDA is an example of a Bayesian topic model. In this, observations (e.g., words) are collected into documents, and each word's presence is attributable to one of the document's topics. Each document will contain a small number of topics.
    - We analyzed the topics found and inspected to see if any patterns or themes appeared.

Sentiment Analysis Tools

We leveraged two sentiment analysis tools to classify the dataset and compare their results.

1. **VADER** (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool developed by researchers at the Georgia Institute of Technology. VADER is specifically attuned to sentiments expressed in social media and works well on texts from other domains. VADER performs best on raw text (with minimal cleaning, without tokenization or removing stopwords). So, for this section of the experiment, we kept capitalization, punctuation (like exclamation marks) and emojis as these can convey emotions or sentiments. We created a bar chart to visualize the VADER sentiment classification distribution.
2. **TextBlob** is a Python library for processing textual data. It provides a consistent API for diving into common natural language processing (NLP) tasks including sentiment analysis. TextBlob provides a polarity score (ranging from -1 to 1) and a subjectivity score, but most commonly we focus on polarity. We can classify the sentiment based on polarity thresholds like VADER. TextBlob tends to be more conservative and often classifies text as neutral if polarity is close to zero. Unlike VADER, TextBlob does not take capitalization, punctuation, or emojis into account explicitly. It uses a lexicon-based approach for polarity. We created a bar chart to visualize the TextBlob sentiment classification distribution.
3. We visualized and compared the results of VADER vs. TextBlob’s sentiment classifications.

Supervised Learning

1. **Data Cleaning**
    - Removed rows with empty tweets
    - Tokenized the tweets: We split up sentences from the tweets into words as a comma separated list of values.
    - Removed stop words: We removed the insignificant words form the tokenised list of words like prepositions, etc.
2. **Sentiment Labelling**
    - The tweets were labelled “Positive”, “Negative”, “Mourning”, and “Neutral” based on the presence of the following words:
        - Mourning: "passed away", "mourning", "sad", "loss", "grieve"
        - Positive: "love", "condolences", "grateful", "respect", "admire", "support"
        - Negative: "hate", "criticize", "anger", "disrespect", "colonial", "resentment"
        - Neutral: If none of the above words matched
3. **Label and Data Encoding**
    - The labels created above were converted into numeric values based on the following logic:
        - "Positive": 0
        - "Negative": 1
        - "Neutral": 2
        - "Mourning": 3
    - The cleaned tweet data was converted into numerical vectors.
4. **Train/Test Split**
    - The data was split into two sets: training set and test set with an 80:20 ratio respectively.
5. **Training the Logistic Regression Model**
    - The logistic regression model was trained using the training set.

Performance Metrics

Unsupervised Learning

Evaluating models and methods without labeled data is inherently challenging. Since we do not have any "ground truth" labels to compare the model's predictions to. However, there are several strategies and indirect metrics we employed to gauge the quality and utility of our results.

Silhouette Scores

- We used silhouette scores to evaluate the K-Means clustering model and find the optimal value of k.
- The silhouette score measures how similar a data point is to its own cluster compared to other clusters. It ranges from -1 to +1, where a higher score indicates better clustering.
- We tested a range of k values from 2 to 11 and plotted a chart to visualize the which k yielded the highest silhouette score.

Evaluating LDA with Topic Coherence Metrics

- Metrics like UMass, UCI, or NPMI topic coherence scores measure how frequently top terms of a topic co-occur and are semantically related, providing an indirect measure of topic quality.

Sentiment Analysis Tools

Evaluating VADER and TextBlob's Performance with Human Inspection

- We randomly sampled 5 tweets from the dataset and manually added sentiment labels based on our judgment (human evaluation).
- Then we compared VADER and TextBlob's predicted classifications with our own labels to check the accuracy and compare the results.

Supervised Learning

The supervised logistic regression model was evaluated using the test set. A confusion matrix was made to visualize the results. We considered the following metrics:

- Accuracy
- Precision
- Recall
- F-Score

Results

Exploratory Data Analysis (EDA)

To understand the dataset better, we identified the top 10 most frequent words in the processed tweets. This revealed that most of the words are not very diverse and do not inherently carry sentiment, which made sentiment analysis challenging.

word	count
queen	155321
elizabeth	128951
ii	117163
queenelizabethii	34954
majesty	31262
pass	25120
rest	23611
death	22429
family	21346
die	21118

Unsupervised Learning

K-Means Clustering

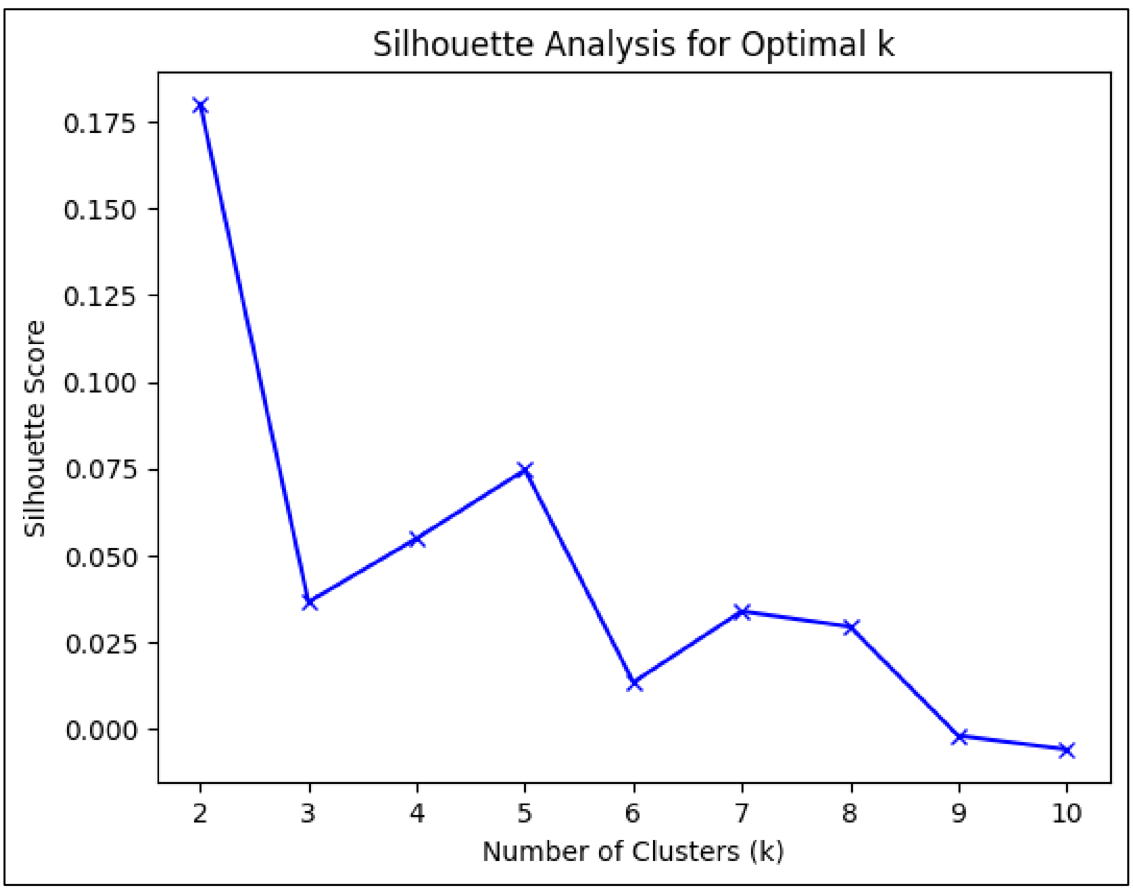
Silhouette Score and Optimal K

The Silhouette Score measures how well each data point fits within its assigned cluster compared to other clusters. Using squared Euclidean distance, the Silhouette Score was 0.0746, when k=5 which is quite low.

This low score reflects the dataset's inherent challenges:

- The vocabulary is repetitive, dominated by terms like "Queen," "Elizabeth," and "RIP."
- Tweets often overlap in themes (e.g., mourning, tributes, and general condolences), making it harder to differentiate distinct clusters.

The optimal number of clusters (k) was found to be 2, suggesting that the dataset contains only two meaningful groupings of content, further indicating that this data is not ideal for clustering. When k= 2 the silhouette score = 0.17, which is still low.



Cluster Interpretations

Cluster 0:

- **Top Words:** ['queen', 'elizabeth', 'ii', 'death', 'rip', 'pass', 'majesty', 'world', 'life', 'die']
- **Analysis:** Focuses on the death of Queen Elizabeth II, with terms like "death," "rip," and "pass." However, frequent mentions of "queen" and "elizabeth" overlap with other clusters, making it less distinct.
- **Coherence:** Mourning theme is present, but the overlap dilutes its meaningfulness.

Cluster 1:

- **Top Words:** ['queenelizabethii', 'queen', 'queenelizabeth', 'rest', 'rip', 'peace', 'death', 'majesty', 'elizabethii', 'world']
- **Analysis:** Centers on tributes and mourning, emphasizing "rest," "rip," and "peace." However, it repeats common terms like "queen" and "queenelizabethii," which overlap with other clusters.
- **Coherence:** Clear focus on tributes, but redundancy weakens its distinctiveness.

Cluster 2:

- **Top Words:** ['queen', 'elizabeth', 'ii', '96', 'die', 'age', 'monarch', 'year', 'britain', '70']
- **Analysis:** Highlights the Queen's age, reign, and legacy, with terms like "96," "monarch," and "year." Despite this focus, repeated words like "queen" and "elizabeth" reduce its uniqueness.
- **Coherence:** Somewhat coherent but overlaps with mourning-related clusters.

Cluster 3:

- **Top Words:** ['queen', 'elizabeth', 'ii', 'family', 'majesty', 'royal', 'pass', 'condolence', 'sadden', 'thoughts']
- **Analysis:** Emphasizes condolences and thoughts for the royal family, using terms like "family," "condolence," and "sadden." The recurring use of "queen" and "elizabeth" again reduces its distinctiveness.
- **Coherence:** Coherent in sentiment, but overlaps with other mourning clusters.

Cluster 4:

- **Top Words:** ['queen', 'elizabeth', 'rest', 'ii', 'peace', 'may', 'majesty', 'queenelizabeth', 'pass', 'queenelizabethii']
- **Analysis:** Focuses on peaceful tributes, with words like "rest," "peace," and "may." It is very similar to Cluster 1, with significant overlap in repeated terms like "queen" and "queenelizabeth."
- **Coherence:** Coherent but lacks originality due to redundancy with other clusters.

The clustering analysis suggests that while the themes of mourning and tribute are present, the clusters tend to overlap significantly, focusing mostly on variations of the same topic—the death of Queen Elizabeth II. The lack of clear differentiation between clusters indicates that the clustering could be improved by reducing the redundancy of these words.

LDA (Topic Modeling)

Coherence Scores

- **UMass Coherence Score: -3.95**
  UMass is based on the co-occurrence of words in the same document (bag-of-words approach).
  Scale: UMass coherence scores typically range from -14 (poor coherence) to 0 (excellent coherence). Higher scores (closer to 0) indicate better topic quality.
  Interpretation:
    - The UMass score indicates moderate coherence. This suggests that while some topics capture related words, others may contain loosely associated or unrelated terms.
    - Since UMass is based on document-level word co-occurrence, the relatively low score may reflect the informal and diverse nature of tweets, where vocabulary and structure vary significantly.

  Dataset Context:
    - Tweets about Queen Elizabeth II's death likely mix emotional and factual content, which can challenge document-level coherence. For example, a tweet mourning the Queen might also reference historical or political themes, blurring topic boundaries.
- **UCI Coherence Score (c_v): 0.462**
  UCI (c_v) measures the word co-occurrence using a sliding window approach across the entire corpus.
  Scale: Scores typically range from 0.0 to 1.0. Higher values indicate more coherent and meaningful topics.
  Interpretation:
    - The UCI score indicates moderate topic coherence. It suggests that the topics contain some meaningful co-occurrences of words across tweets but are not fully optimized.
    - UCI measures coherence across the corpus, so diverse themes (like mourning versus colonial critique) might reduce coherence.

  Dataset Context:
    - The dataset likely contains tweets with highly distinct vocabularies (e.g., "Queen," "condolences," versus "colonialism," "debate"), which makes it harder to group words into semantically tight clusters.

- **NPMI Coherence Score: 0.051**
  NPMI (Normalized Pointwise Mutual Information) measures word association strength within topics.
  Scale: Ranges from -1 (no association) to 1 (strong association). A score of 0 indicates no meaningful relationship between the words.
  Interpretation:
    - The NPMI score is quite low, indicating weak word associations within topics. This suggests that the topics may include words that do not strongly co-occur or are not tightly related in meaning.
    - NPMI is sensitive to rare words and relies heavily on semantic consistency, which informal tweets often lack.

  Dataset Context:
    - Tweets related to Queen Elizabeth II's death might use highly varied and emotionally charged language. For example:
        - Mourning tweets might include words like "loss," "condolences," and "rest in peace."
        - Political critique tweets might use "colonialism," "monarchy," and "oppression."
    - If these distinct vocabularies are mixed within topics, it reduces the coherence score.

Deeper Insights from Coherence Scores

1. **Challenges with Tweets:**
    - Tweets are short and often include informal, fragmented language.
    - They frequently mix sentiments or contexts (e.g. a tweet might mourn the Queen while critiquing the monarchy), leading to overlapping topics.
    - Hashtags, emojis, and unique tweet-specific vocabulary might dilute topic clarity.
2. **Potential Themes and Sentiments in the Dataset:**
    - Clear themes: Topics like "mourning" should ideally show high coherence if preprocessing is effective.
    - Blurred themes: Topics such as "political critique" might overlap with "historical reflections", lowering coherence scores.

Human Inspection/Judgment Analysis of LDA Topics

Topic 0

- **Sample Tweets**:
  - "Uju Anya sends world into a frenzy with shocking tweet about Queen Elizabeth II."
  - "US varsity reacts as Nigerian-born professor wishes Queen Elizabeth II excruciating death."
  - "Dr. Jill Biden was wringing her hands as Joe signed the condolence book for Queen Elizabeth II."
  - "In 1983, Queen Elizabeth II and Prince Philip visited California."
  - "Gorgggg."
- **Analysis**:
  This topic is incoherent, blending controversial opinions, historical references, and unrelated casual remarks. It lacks a clear, interpretable theme, suggesting ineffective clustering for these tweets.

Topic 1

- **Sample Tweets**:
  - "Get the latest stories on the passing of Queen Elizabeth II on Cignal."
  - "Watch live: UK begins a period of mourning following the Queen's passing."
  - "Queen Elizabeth II has died aged 96 at her Scottish home of Balmoral."
  - "A short message: RIP HM Queen Elizabeth II."
  - "A very special lady, one of a kind. RIP Your Majesty Queen Elizabeth II."
- **Analysis**:
  This topic is moderately coherent, focusing on announcements and tributes. While most tweets align with the mourning sentiment, live coverage updates slightly diverge from the central theme.

Topic 2

- **Sample Tweets**:
  - "We join the nation in mourning the death of Her Majesty Queen Elizabeth II."
  - "We are saddened by the passing of Queen Elizabeth II."
  - "The official mourning period for Queen Elizabeth II has begun in Canada."
  - "WomenSTEC mourns the passing of Queen Elizabeth II."
  - "King Charles III has requested 17 days of official mourning."
- **Analysis**:
  This topic is highly coherent, focusing on national and international mourning and condolences. Tweets clearly align with the theme, making this topic meaningful.

Topic 3

- **Sample Tweets**:
  - "Who is this coming from Edom, with his garments stained crimson? Queen Elizabeth II."
  - "I was summoned."
  - "It's the consistency."
  - "It's hot down here ngl."
  - "Farewell to Queen Elizabeth II."
- **Analysis**:
  This topic is incoherent, with a mix of religious allusions, casual remarks, and vague statements. It fails to capture a meaningful theme related to Queen Elizabeth.

Topic 4

- **Sample Tweets**:
  - "Grief is the price we pay for love - Queen Elizabeth II."
  - "Prince Charles named king after Queen Elizabeth II's death."
  - "King Charles III succeeds the throne after the Queen's passing."
  - "Royal family updates: Latest news after Queen Elizabeth II's death."
  - "RIBA president pays tribute to Queen Elizabeth II."
- **Analysis**:
  This topic is coherent, focusing on succession and tributes. The tweets align well with the theme of leadership transition and remembrance.

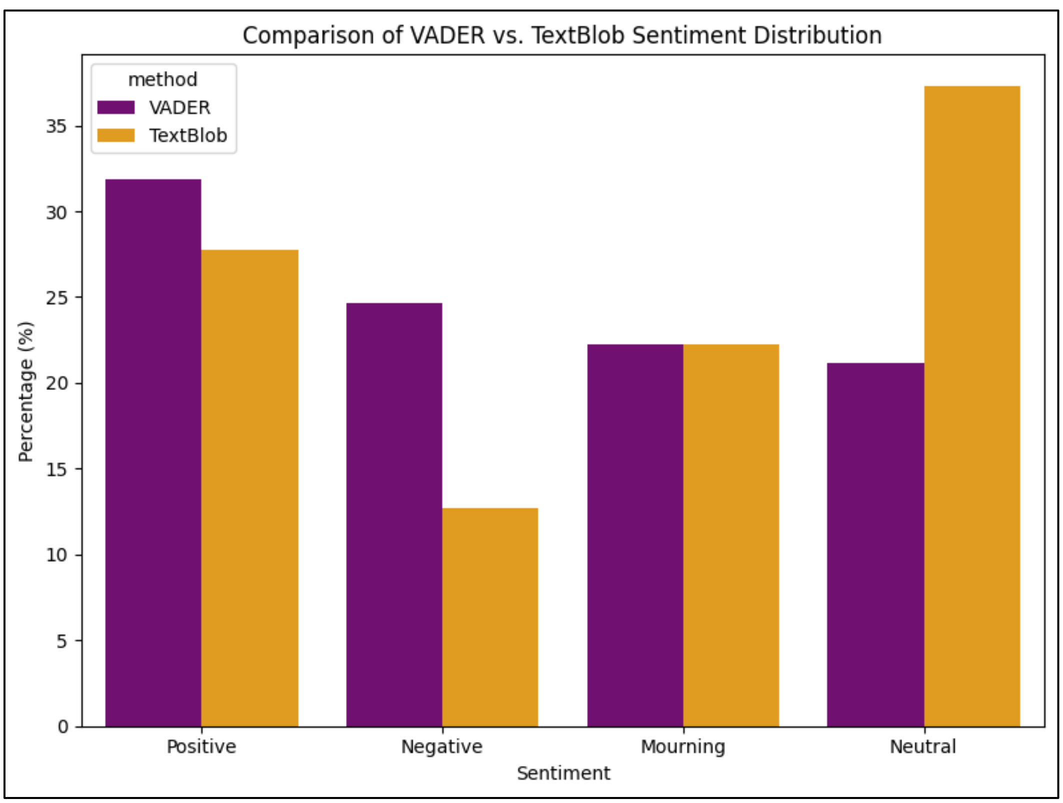
Overall Findings

- **Coherent Topics**: Topics 2 and 4 stand out as coherent and meaningful, capturing themes of mourning and tributes.
- **Incoherent Topics**: Topics 0 and 3 lack clear themes, reflecting the challenges of diverse, informal, and emotional language in tweets.
- **Dataset Challenges**: Repetitive terms like "Queen Elizabeth" and "RIP" appear across multiple topics, reducing the distinctiveness of clusters and highlighting the limitations of the dataset.

VADER and TextBlob

Sentiment Distribution

- **Positive Sentiment:**
  VADER identifies 35% of tweets as positive, which is notably higher than TextBlob's 27%. This aligns with VADER's design, as it is finely tuned for nuanced social media cues, including punctuation, capitalization, and emoticons, which may amplify positive sentiment.
- **Negative Sentiment:**
  VADER also detects more negative tweets (25%) than TextBlob (17%). This could indicate that VADER is more responsive to strong negative cues like words with negative connotations or intensifiers.
- **Neutral Sentiment:**
  TextBlob's significantly higher neutral classification (38% compared to VADER's 21%) suggests that TextBlob might take a more conservative stance in cases of ambiguous sentiment, opting for neutrality.
- **Mourning Sentiment:**
  The "Mourning" category is consistent across both tools (VADER: 22%, TextBlob: 22%). However, this is due to a custom rule we implemented for VADER, detecting mourning-related keywords.



Heuristic Checks

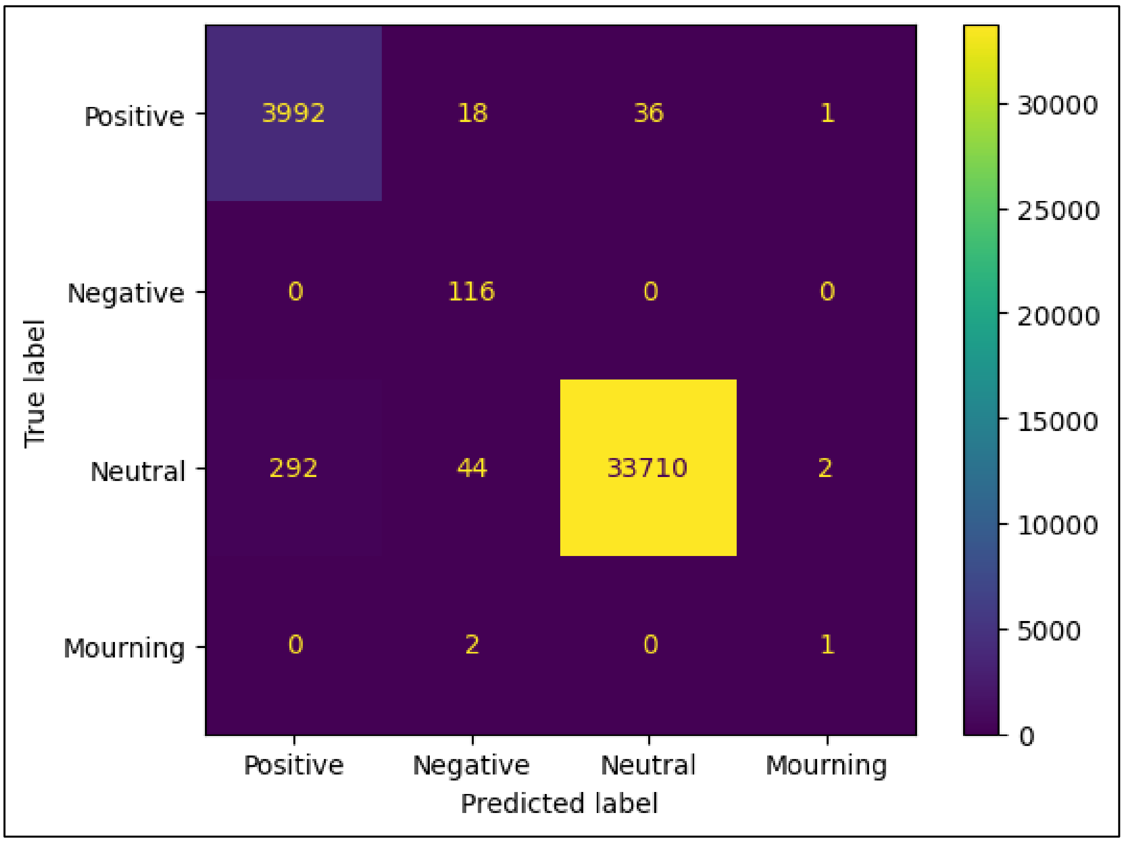
To assess the effectiveness of VADER and TextBlob for sentiment classification, we manually labeled and reviewed a sample of five tweets and compared the sentiment results from both tools with the expected sentiment.

Tweet	Expected Sentiment	VADER Sentiment	TextBlob Sentiment	VADER Match	TextBlob Match
"This profile will now enter a 10-day period of mourning in tribute of Queen Elizabeth II."	Mourning	Mourning	Mourning	✅	✅
"Black Twitter completely unbothered by Queen Elizabeth II's death."	Negative	Negative	Negative	✅	✅
"Top story UK and world react to the death of Queen Elizabeth II."	Neutral	Negative	Positive	❌	❌
"Queen Elizabeth II, thank you for your dedication to public service, an inspiration to us all."	Positive	Positive	Neutral	✅	❌
"Live condolences pour in from across the globe for Queen Elizabeth II's demise."	Mourning	Mourning	Mourning	✅	✅

- VADER performed better overall, matching the expected sentiment for 4 out of 5 tweets, whereas TextBlob only matched 3 out of 5.
- VADER seems more sensitive to the emotionally charged language and nuances in social media content, such as mourning-related expressions.
- TextBlob appears to be more conservative in its classifications, especially when sentiment is ambiguous, and struggles with context-specific nuances.

Supervised Learning

After analyzing the poor-to-mixed results of clustering, LDA, VADER and TextBlob, we decided to also explore a supervised learning approach where we used a logistic regression model. We used very simple heuristics to label the data just by checking for the presence of certain keywords. While this labeling approach is not sophisticated or totally accurate it provides a workaround for the issues faced during unsupervised learning, as discussed in the previous section. Below is the confusion matrix:




```
Accuracy:  0.9896634741194327
Precision:  0.9906928729173492
F1-Score:  0.989961238295044
```

The supervised learning, logistic regression model performed well with near perfect Accuracy, Precision, Recall and F-Scores. However, this is likely due to the simple heuristics we used to label the dataset, typically machine learning models are much more complex.

Conclusion

This project analyzed public sentiment surrounding Queen Elizabeth II's death using unsupervised learning, sentiment analysis tools, and supervised learning. Key findings include:

- **Dataset Challenges**: Repetitive vocabulary and limited diversity hindered clustering and topic modeling, highlighting the need for high-quality datasets.
- **Unsupervised Learning**: K-Means and LDA revealed overlapping clusters and moderate topic coherence due to informal tweet language and thematic overlaps.
- **Sentiment Analysis**: VADER captured nuanced social media cues better than TextBlob but required custom heuristics to detect mourning sentiments effectively.
- **Supervised Learning**: Logistic regression, trained on heuristic-labeled data, achieved high accuracy but reflected the simplicity of its labeling approach.

The study underscores the importance of better datasets and advanced modeling techniques for analyzing complex social media sentiment. Future work could enhance sentiment detection with domain-specific tools and explore temporal sentiment trends.

References

Chintalapudi, N., Battineni, G., & Amenta, F. (2021). Sentimental Analysis of COVID-19 Tweets Using Deep Learning Models. *Infectious Disease Reports*, *13*(2), 329-339. https://doi.org/10.3390/idr13020032

Adams, Emmanuel Jibb & Hassan, Joan. (2024). Assessment of public comments on the death of Queen Elizabeth II: A focus on twitter. Journal of Language, Literature, Social and Cultural Studies. 2. 170-185. 10.58881/jllscs.v2i3.90.