Analysing and modeling of Election Narratives

Gaurav Dhande[1], Munesh Kumar[1], Nivedita Menon[1], Sharath Devanand[1], Shweta Kakade[1], and Swapnanil Ghosh[1]

[1]Affiliation, School of Computer Science, University of Sheffield, Sheffield,UK
[*]gbdhande1, mkumar3, nmenon2, skakade1, sghosh6, sdevanand1 @sheffield.ac.uk

ABSTRACT

Executive Summary

This study delves into the development of a robust classifier capable of assigning both overarching super-narratives and granular thematic narratives to tweets. Our research leverages a rich dataset of tweets collected during the UK 2019 elections, but the methodology and findings hold broader applicability for automated tweet classification across diverse social and political contexts.

The core innovation of our approach lies in the integration of a confidence scoring system alongside the assigned narratives. This transparency mechanism empowers users to evaluate the model's certainty in its classifications, fostering trust and reliability in the analysis process. Employing meticulous data preprocessing techniques, including tokenisation and manual annotation, we meticulously construct a robust training dataset. We then embark on a systematic exploration of various preprocessing techniques and classification algorithms, aiming to optimise the classifier's ability to accurately assign super-narratives and narratives to tweets within the dataset.

Drawing upon insights gleaned from computational linguistics and machine learning literature, we meticulously identify and evaluate strategies to enhance the classifier's effectiveness in accurately categorising tweets. Furthermore, we delve into a detailed analysis of the distribution of assigned labels within the dataset, alongside a rigorous examination of the classifier's generated confidence scores. This comprehensive analysis allows us to assess the model's strengths and weaknesses, ultimately leading to further refinement and improvement.

Through this research, we contribute not only to the advancement of methods for tweet classification, but also offer valuable insights into the challenges and opportunities inherent in automated content analysis on social media platforms. The findings presented in this report hold significant implications for both academic research and practical applications in the field of social media analysis and information processing.

Introduction

In an age where social media has become a principal platform for disseminating information, narratives, and opinions, understanding the nature of these digital conversations is crucial. This project focuses on developing a robust classifier capable of accurately identifying both overarching super-narratives and detailed thematic narratives within tweets, using data collected from the UK General Elections in 2019. While centered around this specific political event, the methodology and findings have broader implications for automated classification across diverse social and political contexts.

The UK 2019 General Elections highlighted the rapidly shifting landscape of public discourse on platforms like Twitter. Polarization, misinformation, and strategic political messaging were prevalent, reflecting the changing dynamics of political communication in the digital era. This made the election period an ideal case study to analyze the challenges and nuances inherent in identifying and categorizing narratives. Moreover, as public sentiment shifts and misinformation spreads, developing reliable systems for content analysis is crucial for researchers, policymakers, and organizations.

The primary challenge lies in capturing the intricate nature of narratives within limited tweet text, often constrained by the platform's character limit. This project addresses this challenge by employing a classifier that incorporates a confidence scoring system alongside the assigned narratives. The confidence scores empower users to assess the model's certainty, fostering trust in the analysis process. By providing both transparency and reliability, the system serves as a valuable tool for social media analysts and practitioners.

The methodology of the project involves meticulous data preprocessing, including tokenization, cleansing, and manual annotation, to build a comprehensive training dataset. This phase is crucial for removing irrelevant, redundant, or spam

content and ensuring the integrity of the data. The research then systematically explores various preprocessing techniques and classification algorithms to optimize the classifier's performance, including advanced language models like BERT, RoBERTa, and LLAMA3. Additionally, embeddings such as GloVe are leveraged to capture nuanced semantic relationships, ensuring that the classifier can effectively recognize contextual patterns and themes.

To achieve robust classification, the project utilizes several machine learning models, including Naive Bayes, Logistic Regression, Random Forest, Linear SVM, and Multilayer Perceptron, among others. These models are evaluated through rigorous experimentation, testing, and parameter tuning, providing insight into their strengths and weaknesses across various narrative categories. The classifier's generated confidence scores are then analyzed alongside the assigned labels to assess the model's strengths and weaknesses, allowing further refinement and improvement.

This research provides valuable contributions to the field of social media analysis by advancing automated methods for narrative classification. It offers insights into the challenges and opportunities inherent in analyzing political discourse on platforms like Twitter. The findings presented hold significant implications for both academic research and practical applications in social media analysis, content moderation, and information processing.

By systematically examining the distribution of super-narratives and narratives within the dataset, and rigorously assessing the classifier's accuracy and confidence scores, this research reveals critical trends and patterns. It uncovers the complex nature of public sentiment, political polarization, and misinformation in the digital age, providing a deeper understanding of how narratives shape and influence political outcomes.

Ultimately, this project seeks to establish a framework that not only enhances the classification of social media narratives but also serves as a comprehensive resource for navigating the complexities of digital discourse. In doing so, it offers a more informed and nuanced approach to information processing and political analysis, helping stakeholders better understand the rapidly evolving landscape of social media conversations.

Background and literature review

Kotseva[1] examined 58,625 articles from 460 unverified sources over three years to identify and categorize COVID-19 mis/disinformation narratives. This massive dataset provided an expansive view of misinformation trends and narratives, reflecting the global scale and complexity of the pandemic's information landscape. Using NLP approaches, the researchers constructed a hierarchical codebook containing 12 super narratives, 51 narratives, and 44 subnarratives, revealing major themes such as fearmongering, criticism of institutions, and conspiracy theories. Their BERT-based model enabled real-time tracking of how these narratives evolved and spread, offering valuable insights into public perception and behavior during a public health crisis. Articles were collected through the Europe Media Monitor (EMM) system and translated into English, while text clustering was conducted using Term Frequency-Inverse Document Frequency (TF-IDF) and Latent Semantic Analysis (LSA). Two annotators manually organized narratives inductively, and the BERT-based classifier was oriented on 30,000 annotated articles. The findings showed that fearmongering dominated early pandemic discourse but gradually diminished after spring 2020 as the understanding of the virus increased. Criticism of authorities like the EU, WHO, and national governments peaked in mid-2020 and intersected with geopolitically charged narratives moving between pro-China, anti-China, pro-Russia, and anti-Russia compositions. Conspiracy theories remained consistent throughout the pandemic, with occasional spikes, while vaccine-related misinformation surged as vaccines became available, targeting vaccine hesitancy and mandatory vaccination. Despite the comprehensive analysis, limitations included a geographic bias focusing on Western sources, clustering challenges using TF-IDF algorithms, and translation artifacts leading to linguistic misclassification. Kotseva's hierarchical codebook presents a structured understanding of recurring mis/disinformation themes, useful for pinpointing patterns and directing classification and counter-narrative strategies.

Santana[2] delivers a comprehensive summary of narrative extraction, defining narratives as sequences involving actors and events across time and space, highlighting linguistic and computational challenges in learning and automating narrative extraction using NLP techniques. The survey establishes a foundational framework for future research by providing a framework encompassing events, actors, relationships, and temporal/spatial data. Santana emphasizes the challenges inherent in manually annotating texts to train NLP models, considering multi-layered semantic representation and annotation schemes. The five-step extraction pipeline—pre-processing, identification, linking components, representation, and evaluation—addresses foundational issues. Pre-processing and parsing are essential for tokenization, normalization, and syntactic parsing, but these steps are complicated by informal or multilingual text. Critical components like events, participants, time, and space require advanced NLP techniques for precise extraction, while linking components demands temporal reasoning and entity relation extraction to ensure coherent interpretation of the narrative structure. Despite significant advancements, the field still lacks a standard evaluation framework, and invariant metrics are required for cross-analysis comparability. Cross-domain applicability is also challenging due to varying annotation schemes and narrative complexity across domains. Automated processes struggle with diverse genres, informal languages, and long-form texts like novels, while temporal and spatial ambiguity persists in handling vague or implicit references. This survey aligns with the project's purposes, clearly specifying the key components

and challenges in narrative extraction. Santana's five-step extraction pipeline provides a clear framework for categorizing and linking narrative elements, directly informing methodologies for detecting and classifying disinformation .

Metilli, Bartalesi, and Meghini[3] designed a system for extracting formal narratives from text, emphasizing that narratives consist of events set in space and time, interlinked via semantic relationships. They identified eight technical conditions necessary for narrative extraction, including event detection, classification, named entity recognition, and temporal entity extraction. Their neural network-based approach, using a recurrent neural network with bidirectional Long Short-Term Memory (LSTM) architecture, detects and classifies events. The model was trained on the ACE 2005 corpus and manually annotated Wikipedia biographies, ensuring a robust and diverse training dataset. The model's performance was evaluated on the biography of Florentine poet Dante Alighieri, focusing on 12 event classes relevant to his life. This evaluation provided a real-world use case demonstrating the model's effectiveness, achieving an F1 score of 73.0 for event detection and 70.3 for event classification. The authors also devised an annotation tool to build and export training datasets in JSON format, achieving an inter-annotator agreement of 0.86, demonstrating the tool's reliability. A web interface was presented for event detection, entity linking, and manual editing, streamlining narrative construction by simplifying the identification, linking, and representation of narrative elements. However, the system is limited to 12 event classes, relevant only to Dante's life, and struggles with the automated linking of events to branches. It requires further development to establish reliable semantic relationships and doesn't fully leverage external knowledge bases for disambiguation. Despite these gaps, the proposed system aligns well with the project's methodology, with neural networks providing efficient event detection and classification. The challenges of entity and event linking resonate with the difficulties in formalizing disinformation narratives, and the web-based visualization interface could inspire similar tools for identifying and analyzing mis/disinformation.

Panizio[4] documented across-the-board disinformation during the 2023 national elections across Europe, investigated through over 900 fact-checking articles. The European Digital Media Observatory's Task Force determined narratives that undermined democratic processes and public confidence in elections, providing a unique insight into the disinformation landscape around electoral processes. Key disinformation themes included voter fraud, foreign influence, and unfair practices. Data collection involved a systematic study of fact-checking articles before and after each election, carefully identifying false narratives and stories. Narratives were categorized and color-coded by topics like electoral processes, geopolitical issues, and social themes. Panizio's analysis identified common disinformation themes, including overall claims of electoral fraud, specifically allegations of vote tampering and misinformation regarding the voting process. Narratives varied widely by local context, impacted by factors such as the war in Ukraine, economic conditions, climate change, and social issues like immigration and gender. However, the scope of analysis focused primarily on EU countries, potentially overlooking broader global trends, and the reliance on fact-checking organizations may miss less-publicized narratives. Despite these limitations, this report provides a comprehensive summary of election-related disinformation, delivering insights into strategies used to influence public opinion. Understanding the thematic classification of narratives is essential for the project's analysis of election-related mis/disinformation narratives and the identification of emerging trends.

Ahmad's[5] working paper discusses how online misinformation is financially supported through advertising, with a particular focus on the functions of advertisers and digital ad platforms. Descriptive analysis and survey-based experimentation explored how misinformation websites secure advertising revenue and how consumer behavior shifts when made aware of this association. The study also examined how decision-makers reacted to discovering that their companies advertised on misinformation outlets. Data was collected from NewsGuard, the Global Disinformation Index (GDI), and Oracle's Moat Pro platform to analyze advertising behavior across 10,310 websites from 2019 to 2021. In a consumer survey experiment, a representative sample of U.S. internet users (4,000 participants) was surveyed about their response to companies advertising on misinformation websites. A decision-maker survey explored executives and managers' awareness and preferences regarding their companies' advertising practices. The study found that 44% of advertisers from a dataset of 42,595 appeared on misinformation websites, and those using digital ad platforms were ten times more likely to have their ads appear on misinformation outlets. When consumers learned about companies advertising on misinformation sites, many switched their brand preferences. Many executives were unaware that their companies' ads appeared on these sites, but after learning, they strongly preferred to avoid such associations. However, the study primarily focused on websites in English-speaking regions, which may not fully capture the global advertising ecosystem. Additionally, survey participants may vary in their perception of misinformation websites, affecting the generalisability of the results. This study addresses the project's goals of narrative detection and understanding the financial incentives behind disinformation. It emphasizes how advertising platforms amplify disinformation by providing monetization channels, underscoring the importance of understanding these dynamics to refine narrative detection strategies.

The paper by Oates, Lee, and Knickerbocker[6] presents a data-driven approach to detect and analyze Russian disinformation narratives and trace their penetration into the U.S. media ecosystem. The authors focused on narratives related to accusations of Nazism against the Ukrainian Azov Battalion and false flag operations. By leveraging the VAST-OSINT (Open Source Intelligence) system, they provided insights into how Russian disinformation "supply chains" propagate propaganda across various media platforms. The VAST-OSINT system collected and categorized relevant online content from over 3 billion

URLs, while NLP and network analysis were used to identify linguistic patterns consistent with Russian propaganda narratives. Case studies focused on spreading the "Ukrainian Nazis" narrative associated with the Azov Battalion and claims of U.S. false flag operations. The analysis identified 93 unique URLs promoting the narrative that the Ukrainian Azov Battalion is a neo-Nazi group, primarily on far-right U.S. conspiracy websites. The study also uncovered instances where Russian propaganda claimed the U.S. was planning false flag operations to justify military action. The authors observed that right-wing websites amplified Russian narratives, while mainstream U.S. media showed minimal engagement, though niche conspiracy platforms provided a substantial echo chamber. The analysis focused on English-language and right-wing U.S. media, potentially overlooking narratives in other languages or political contexts, while NLP challenges made identifying exact sources difficult due to deliberate obfuscation. Despite these limitations, the methodology and findings align directly with the project's goals, demonstrating how strategic disinformation supply chains can be mapped to gain crucial insights. Understanding which narratives resonate in different environments is essential for detecting and countering strategic disinformation messaging .

The "RESIST 2 Counter-Disinformation Toolkit," prepared by the UK government[7], provides a systematic, evidence-based framework for recognizing and fighting disinformation. It is an update to the original RESIST toolkit, reflecting new realities in the evolving information environment. The framework offers methods for recognizing and categorizing disinformation, establishing early warning systems, and formulating strategic communications. The RESIST 2 framework stands for Recognize, Early Warning, Situational Insight, Impact Analysis, Strategic Communication, and Tracking Effectiveness. Disinformation is categorized into misinformation (without intent to deceive), disinformation (with intent), and misuse of accurate information. Case studies from around the world demonstrate the principles in action, emphasizing early warning through monitoring tools to detect disinformation threats promptly. The framework encourages narrative recognition, highlighting the importance of recognizing recurring themes and symbolism across campaigns. Strategic communications suggest proactive and reactive strategies like pre-bunking and counter-narratives, while impact assessment provides techniques for analyzing disinformation's effects on policy, reputation, and trust. The toolkit's effectiveness depends on consistent implementation across organizations, while continuous monitoring remains resource-intensive yet crucial for early warning. The RESIST 2 toolkit aligns with the project's goal of understanding and countering disinformation narratives. Its step-by-step framework can guide systematic identification, analysis, and strategic response to harmful narratives, while emphasizing proactive communication strategies, like pre-bunking, aligns with the project's goals.

Damian Milewski's paper analyzes the global spread of disinformation and propaganda during the COVID-19 pandemic, focusing specifically on China, Russia, and the USA. The author examines how state or state-sponsored entities manipulate existing misinformation to craft narratives that serve their strategic objectives. Data collection and analysis are based on press and government information using an open-source approach, emphasizing analysis, synthesis, and deduction. Case studies highlight specific narratives and campaigns from China, Russia, and the USA, emphasizing the forms, methods, and tools employed. The pandemic created ideal conditions for information warfare due to widespread public fear and confusion. Disinformation campaigns exploited existing divisions and leveraged trusted "super-spreaders." China's narrative emphasized that COVID-19 originated outside China while positioning the nation as a global leader in pandemic management, utilising the "three wars" concept to influence perception through psychological, media, and legal warfare. Russia's narrative stoked fear and distrust through conspiracy theories about Western governments and the USA, while the "Gerasimov Doctrine" underpins Russia's hybrid warfare strategy, blurring the lines between peace and war. The USA aimed primarily to counter Chinese and Russian disinformation, but its messaging was inconsistent and lacked a comprehensive strategy. Open-source reliance potentially limits comprehensive insights into covert disinformation efforts, while identifying sources remains challenging due to deliberate obfuscation. Nonetheless, Milewski's paper provides a detailed understanding of global disinformation campaigns, crucial for the project's objectives. Analyzing geopolitical actors' use of disinformation helps inform the project's strategy for detecting, countering, and educating against these narratives.

Research question

This study focuses on developing a robust model tailored to categorise specific types of disinformation narratives within the UK 2019 elections tweet dataset. The primary objective is to design a framework capable of assigning both super-narratives (broad labels) and narratives (narrowed labels) to tweets, along with confidence scores based on a pre-defined metric (e.g., margin of classification) that indicate the model's certainty in the assigned labels.

Expanding upon this foundation, our secondary research questions explore comparative analyses between automated and manual annotation processes. By leveraging insights from existing literature on annotation methodologies and data preprocessing techniques, we aim to evaluate the effectiveness of different approaches in accurately categorising tweets within annotated datasets. This examination will provide valuable insights into the strengths and limitations of automated versus manual annotation methods, contributing to the broader discourse on data labelling practices in computational linguistics research.

1. **Objective 1**: Improve the model's label assignment accuracy for disinformation narratives through the exploration of various preprocessing techniques, such as named entity recognition (NER), aimed at identifying fabricated stories, manipulated statistics, and other specific types of disinformation narratives.

2. **Objective 2**: Identify the most effective combination of embedding methodologies and classification algorithms for this specific dataset, leading to improved accuracy in categorising disinformation narratives.

3. **Objective 3**: Conduct data analysis on human annotated tweets to extract useful insights into the characteristics and distribution of narratives, informing further model refinement and optimization strategies.

4. **Objective 4**: Address the imbalance in the dataset through techniques such as oversampling of minority classes or adjusting class weights, ensuring that the classifier maintains high performance across all classes despite variations in data distribution.

Moreover, our research distinguishes itself by focusing on the development of a confidence scoring system alongside the super-narrative and narrative labels. This innovation allows users to gauge the model's certainty in its classifications, thereby enhancing the transparency and reliability of the labelling process.

Drawing from insights in the literature on narrative extraction and computational linguistics, our investigation aims to identify additional features and strategies to enhance the model's effectiveness in identifying and categorising specific types of narratives within tweets. By examining existing research on narrative structure and discourse analysis, we seek to refine the model's capabilities in accurately capturing the nuances of textual narratives, thereby improving its overall performance in classifying narratives.

Methodology

The methodology section details the comprehensive procedures and analytical approaches utilized to model Twitter narratives and super narratives from the United Kingdom's General Elections in November and December 2019. This research aims to decipher broad conversation themes and their respective sub themes as articulated through public tweets. The sections below describe the data collection, structuring, annotation process, and analytical techniques applied.

Data Aggregation

Data for this study was collected via the Twitter API, targeting tweets from the period of the UK General Elections 2019. Initially, a "silver annotation" process was implemented with the intention of creating a robust training set. This method involved selecting a sample of tweets which were then pre-annotated using simpler, semi-automated methods in hopes that they could serve as a preliminary training dataset. The aim was to facilitate the development of a more sophisticated analysis model, particularly leveraging Large Language Models (LLMs) to automate the categorization of narratives and supernarratives within tweets.

However, the silver annotation approach encountered significant challenges. Despite the potential for efficiency gains, this method proved ineffective for several reasons. First, the computational demands of running LLMs were high; the infrastructure required to process data continuously and reliably was substantial. Second, the output from these models was often inconsistent, with issues such as multiple labels being assigned to a single tweet, which complicated the categorization process rather than simplifying it. Additionally, these models sometimes behaved erroneously, misinterpreting tweet contexts or failing to recognize subtle nuances in language that are critical for accurate narrative identification.

Due to these complications, a strategic pivot was necessary. The study shifted towards establishing specific inclusion and exclusion criteria that emphasized the necessity for tweets to be standalone and convey complete thoughts. This change ensured that each tweet could be understood and analyzed independently of any conversational threads, thereby maintaining a clear analytical focus on primary narratives.

Tweets were extracted as raw text, along with metadata such as timestamps and anonymised user information, strictly adhering to privacy considerations. The collection parameters were meticulously set to include keywords and phrases closely associated with the election, such as "UK elections," "voting," "political parties," along with terms relevant to major campaign issues. Following collection, the dataset underwent a rigorous cleansing process to remove non-English tweets and clear spam, utilizing both automated scripts and manual reviews to ensure the integrity and relevance of the data for analytical purposes. This approach significantly improved the quality of the data used in the study, enabling a more accurate and reliable analysis of election-related narratives on Twitter.

Data Structure

The structured dataset utilized in this study comprises individual tweets sourced from Twitter's API, each represented as a JSON object. The primary focus during data extraction was to capture each tweet along with a unique key that could be referenced back to the original tweet if required at a later stage of analysis. To achieve this, the following root-level JSON objects were selected:

1. `id`:
   An Int64 integer representing the unique identifier for each tweet. This identifier is greater than 53 bits, and while some programming languages may encounter difficulties interpreting it, using a signed 64-bit integer ensures safe storage.

2. `text`:
   A string field containing the actual UTF-8 text of the tweet. This text captures the content of the status update, including any characters or symbols used by the user. It's essential for understanding the context and content of each tweet during analysis.

3. `in_reply_to_status_id`:
   An Int64 integer field, nullable, indicating the original tweet's ID if the represented tweet is a reply. This field facilitates the tracking of tweet threads and conversations, providing insights into the interactivity and engagement levels of users. The dataset's structure prioritizes the text content of the tweets, augmented by metadata such as timestamps and anonymised user identifiers. This combination of textual content and metadata enables a comprehensive analysis of the narratives and super narratives surrounding the UK General Elections.

A complete and exhaustive JSON structure of the tweet can be obtained from Twitter's official Data Dictionary Standard

To facilitate nuanced analysis, a hierarchical taxonomy was developed, outlined in a comprehensive codebook accompanying the dataset. This codebook serves as a guide for annotators and researchers, delineating ten super narratives deemed pivotal in the context of the elections. These super narratives encompass a broad range of topics, including "Gender-Related," "Religion Related," "Ethnicity Related," and "Political Hate," among others.

Within each super narrative, multiple narratives are defined, providing a fine-grained classification schema to capture the diverse discourse surrounding the elections. For example, under the super narrative "Political Hate," narratives may include "Pro-right," "Pro-left," "Anti-left," and "Anti-right." The relationship between super narratives and narratives is explicitly documented in the codebook, ensuring consistency in annotation and subsequent analysis. This structured approach enables researchers to navigate the complexities of election discourse effectively and draw meaningful insights from the dataset.

Data Annotation

During the annotation process, each of the 1200 tweets in the dataset underwent individual evaluation by six human annotators, each responsible for labeling 200 tweets. This distributed approach not only facilitated the efficient annotation of a large volume of data but also introduced redundancy, enabling inter-annotator agreement analysis to assess the consistency and reliability of the annotations.

To validate the reliability of the annotations, a separate test set comprising 100-200 tweets was selected and annotated by at least two annotators independently. This validation step served multiple purposes. Firstly, it allowed for the identification of potential discrepancies or disagreements between annotators, highlighting areas where additional clarity or guidance may be required in the annotation instructions. High inter-annotator agreement indicates a consistent and reliable annotation process, bolstering confidence in the quality of the labeled dataset.

During the annotation process, annotators were guided by a predefined taxonomy outlined in the codebook. This taxonomy delineated ten super narratives, each encompassing multiple narratives, providing a hierarchical framework for categorizing the content of the tweets. The codebook served as a reference guide, ensuring consistency and uniformity in the assignment of labels across annotators.

In addition to assigning super narratives and corresponding narratives, annotators provided a confidence score ranging from 1 to 5, indicating their level of certainty in the assigned labels. This confidence scoring mechanism offered insights into the annotators' subjective assessments of the tweet content and their confidence in applying the predefined taxonomy. If a tweet received a confidence score of 3 or lower, the same annotator provided a secondary annotation. This iterative process of secondary review enhanced the robustness of the annotation process, mitigating the potential for misclassification or errors and ensuring the accuracy and reliability of the annotated dataset.

Overall, the annotation methodology employed rigorous quality control measures to uphold the integrity of the labeled dataset, fostering confidence in the subsequent analysis and interpretation of the election discourse on Twitter.

Data Analysis

Following the gold annotations a through analysis was carried out to gain insights regarding the distribution of super narratives and narratives.

The analysis revealed a significant dominance of tweets classified under "No narrative provided," accounting for 545 instances. This category includes tweets that lacked a discernible narrative or were strictly informational, suggesting that a substantial portion of the discourse was centered around neutral information sharing or non-committal commentary. This finding indicates that while political narratives were prevalent, there was also a high level of general communication or possibly disengagement from specific political discussions.

"Political hate and polarisation" and "Distrust in institutions" emerged as the second and third most prominent narratives with 326 and 175 mentions respectively. This narrative underscores the deeply divided nature of public opinion during the election, reflecting strong biases, opposition, or discontent with opposing parties or ideologies coupled with a significant level of skepticism and lack of confidence in traditional institutions, possibly including governmental bodies, the electoral system, or other societal structures.

Less frequent but still significant were narratives such as "Anti-Elites," "Geopolitics," and various identity-related themes like ethnicity, gender, migration, and religion. These discussions, while not as widespread, highlight the diversity of issues that can influence voter behavior and election outcomes.

Model Selection

Embeddings

BERT is a pre-trained language model developed by Google. It's bidirectional, meaning it considers context from both the left and right side of a word in a sentence. BERT learns contextual representations of words by training on large amounts of text data. It consists of multiple Transformer layers, allowing it to capture complex linguistic patterns and relationships. BERT embeddings can be fine-tuned for specific downstream tasks like classification, named entity recognition, and question answering.

RoBERTa is an extension of BERT, developed by Facebook AI. It addresses some of the limitations of BERT by training on more data, for longer periods, with larger batch sizes, and with dynamic masking patterns. RoBERTa achieves better performance on various natural language understanding tasks by refining BERT's training methodology and hyperparameters.

LLAMA3 is a contextual language model developed by Salesforce Research. It's trained on diverse datasets and tasks simultaneously, allowing it to generalize better across different domains and tasks. LLAMA uses a mixture of unsupervised pre-training and supervised fine-tuning to adapt to specific tasks efficiently. It achieves state-of-the-art performance on various natural language processing tasks, including text classification, sequence labeling, and text generation.

GloVe is an unsupervised learning algorithm for obtaining vector representations (embeddings) for words. Unlike BERT and its variants, GloVe does not capture contextual information. Instead, it leverages global word co-occurrence statistics from large text corpora to learn word embeddings. GloVe embeddings represent the semantic relationships between words based on their co-occurrence probabilities. These embeddings are widely used in tasks like word similarity calculation, language translation, and sentiment analysis, where capturing global word semantics is crucial. GloVe embeddings are typically pre-trained and then fine-tuned for specific downstream tasks.

Elementary Machine Learning Models

An exploratory analysis of various machine learning models are implemented to understand the advantages and disadvantages that the data has on each of the models. Below is the list of the models that were utilized

- Multinomial Naive Bayes - Multinomial Bayes is a probabilistic classifier based on Bayes' theorem with the assumption of independence between features, often used for text classification tasks with discrete features.
- Logistic Regression - A linear classification algorithm that predicts the probability of a binary outcome based on input features, widely used for binary classification tasks.
- Random Forest - An ensemble learning method that constructs a multitude of decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees.
- Linear SVM - A linear classifier that separates classes by finding the hyperplane that maximizes the margin between classes in feature space.
- Multilayer Perceptron - A type of artificial neural network composed of multiple layers of nodes, capable of learning complex patterns in data and widely used for classification and regression tasks.

- Decision Tree - A tree-like structure where internal nodes represent features, branches represent decisions, and leaf nodes represent the outcome, used for classification and regression tasks.
- K-Nearest Neighbors - A non-parametric method used for classification and regression, where the output is based on the majority (or average) of the 'k' nearest data points in the feature space.
- AdaBoost - An ensemble learning method that combines multiple weak classifiers to form a strong classifier, with each classifier giving more weight to the misclassified points from the previous one.
- Gradient Boosting - An ensemble learning technique that builds decision trees sequentially, where each tree corrects errors made by the previous one, producing a strong predictive model.
- XGBoost - An optimized implementation of gradient boosting designed for speed and performance, widely used in machine learning competitions and production systems.
- Voting Classifier - An ensemble method that combines multiple machine learning models and predicts the class based on the majority vote of all models.
- Stacking Classifier - An ensemble learning technique that combines multiple classification or regression models via a meta-classifier or a meta-regressor, often achieving higher performance than individual models.

Hyperparameter Tuning

Hyperparameter tuning is a crucial step in optimizing the performance of machine learning models. For Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Stacking Classifier, this process involves adjusting various parameters to enhance predictive accuracy and generalization.

SVM's hyperparameters, including the choice of kernel and regularization parameter, are tuned to optimize its performance. SVM is advantageous for its ability to handle high-dimensional data and nonlinear decision boundaries efficiently. By tuning SVM, we aim to find the optimal settings that balance model complexity and generalization.

MLP's hyperparameters, such as the number of hidden layers, neurons per layer, and activation functions, are adjusted to capture complex nonlinear relationships in the data. MLP offers flexibility and adaptability to diverse problem domains, making it a versatile choice. By tuning MLP, we seek to identify the architecture and hyperparameters that yield the best performance on the validation set.

A stacking Classifier combines the predictions of multiple base models using a meta-learner to improve overall predictive accuracy. Hyperparameter tuning for Stacking involves selecting the base models, their respective hyperparameters, and the meta-learner's parameters. Stacking leverages the strengths of different models while mitigating their individual weaknesses, resulting in potentially superior performance. Through hyperparameter tuning, we aim to optimize the composition of the stacked ensemble for maximum predictive power.

By carefully tuning the hyperparameters of SVM, MLP, and Stacking Classifier, we can enhance their performance on our specific dataset and task. This iterative process involves experimentation with different parameter settings and evaluation of their impact on model performance. Ultimately, hyperparameter tuning plays a vital role in maximizing the effectiveness of these machine learning models in real-world applications.

Binary Chaining

The binary chaining algorithm offers a systematic approach to address the challenges presented by imbalanced datasets, where one class significantly outweighs the other. It relies on logistic regression models and a chaining mechanism to sequentially predict the labels of data points. In this method, each unique label within the dataset is individually fitted to a logistic regression model, starting with the most frequent label.

The chaining process begins by training a logistic regression model on the most prevalent label in the dataset. This model is then utilized to predict whether a given data point belongs to the most frequent class or not. If a data point is predicted to belong to the majority class, it is labeled accordingly. However, if the prediction indicates otherwise, the algorithm moves to the next most frequent label in the dataset and repeats the process. This sequential prediction continues until a data point is assigned a label.

The decision to halt the chaining process and assign a label to a data point hinges on the prediction outcome. If a data point is confidently predicted to belong to a class with high frequency, the chaining process concludes, and the corresponding label is assigned. However, if none of the logistic regression models are able to confidently assign a label to a data point, additional strategies may be employed, such as assigning the majority class label or using ensemble techniques to combine predictions from multiple models.

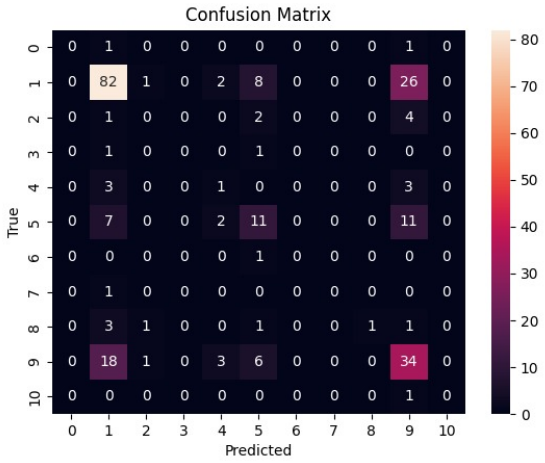

Figure 1. Confusion Matrix for BERT Classifier

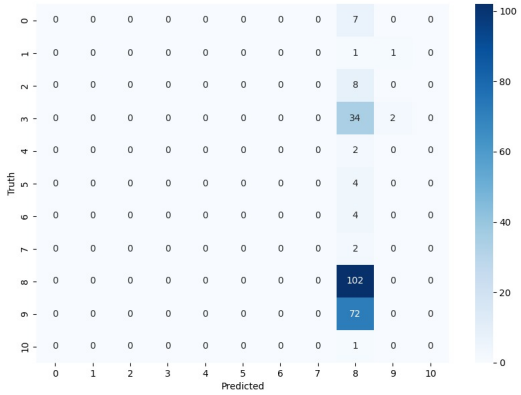

Figure 2. Confusion Matrix for Binary Chaining Classifier

Overall, the binary chaining algorithm provides a structured framework for making predictions on imbalanced datasets. By leveraging logistic regression models and a sequential prediction strategy, it aims to mitigate the effects of class imbalance and make accurate predictions on imbalanced data. This approach can be iteratively refined and adapted to different datasets and problem domains, allowing for improved performance and robustness in handling imbalanced data.

Results

The analysis of the classification reports from the BERT and binary chain classifiers reveals distinct performance characteristics for each. The BERT classifier exhibits a mix of performance metrics across various labels, with some categories achieving moderate success; specifically, label '0' shows a precision of 0.70 and a recall of 0.69, suggesting more effective identification compared to other classes. However, most other classes in the BERT report demonstrate very low or zero precision and recall, indicating struggles with certain classifications or possibly issues related to data representation and class imbalance. In figure 1 you can see the confusion matrix for the BERT classifier.

In contrast, the binary chain classifier displays uniformly poor performance across all categories, with zero scores in precision, recall, and f1-score for the categories listed such as 'Anti-EU' and 'Anti-Elites'. This pervasive under-performance might be attributed to insufficient training data, poor model fit, or the challenges inherent in managing imbalanced datasets.

These findings highlight a significant disparity in the effectiveness of the two models, with neither achieving optimal results across the board. The data suggests that both classifiers could benefit from a reassessment of their training regimes, a more balanced dataset, or a review of model parameters. Additional diagnostics like confusion matrices and ROC curves could also

provide deeper insights into specific areas of weakness and potential strategies for model improvement.

Discussion and conclusions

Silver Annotations
Double Annotations
Manual Annotations
Modelling

References

1. Kotseva, B. *et al.* Trend analysis of covid-19 mis/disinformation narratives–a 3-year study. *Plos one* **18**, e0291423 (2023).
2. Santana, B. *et al.* A survey on narrative extraction from textual data. *Artif. Intell. Rev.* **56**, 8393–8435 (2023).
3. Metilli, D., Bartalesi, V., Meghini, C. *et al.* Steps towards a system to extract formal narratives from text. In *Text2Story@ ECIR*, 53–61 (2019).
4. Disinformation narratives during the 2023 elections in europe.
5. Ahmad, W., Sen, A., Eesley, C. & Brynjolfsson, E. The role of advertisers and platforms in monetizing misinformation: Descriptive and experimental evidence. Tech. Rep., National Bureau of Economic Research (2024).
6. Oates, S., Lee, D. & Knickerbocker, D. Data analysis of russian disinformation supply chains: Finding propaganda in the us media ecosystem in real time. *Oates, Sarah, Doowan Lee, David Knickerbocker* (2022).
7. Resist 2 counter-disinformation toolkit.

Appendices

Super Narrative	Narrative	Description
Anti-EU	EU Economic Skepticism	Narratives focusing on criticism of EU economic policies a their impact.
	Crisis of EU	Discusses perceived crises within the EU, such as financial migration issues.
	EU Political Interference	Criticism of the EU's influence over national sovereignty a policy-making.
	EU Corruption	Accusations of corruption and unethical behavior within E institutions.
Political Hate and Polarization	Pro-far Left	Supports extreme left-wing ideologies and policies.
	Pro-far Right	Advocates for far-right perspectives and policies.
	Anti-Liberal	Opposes liberal ideologies, often emphasizing conservative v ues.
	Anti-Left	Expresses opposition to left-wing politics and ideas.
	Us vs Them	Highlights division and conflict between differing groups ideologies.
Religion-related	Anti-Islam	Narratives that criticize or oppose Islamic religion or Muslim
	Religious-Sexist Narratives	Links religious beliefs with sexist attitudes or policies.
	Anti-vax Narratives	Religion-based opposition to vaccination and medical scienc
	Disease Spreaders	Accusations targeted at religious groups as carriers of disease
	Anti-Semitic Conspiracy Theories	Narratives promoting conspiracy theories targeting Jewish pe ple.
	Interference with State Affairs	Claims of religious groups interfering in secular governme activities.
Gender-related	Language-Related	Discussion of gender-specific language and its societal impac
	LGBTQ+-Related	Focuses on narratives related to LGBTQ+ rights and issues.
	Demographic Narratives	Examines changes in gender demographics and implications f society.
Ethnicity-related	Association to Political Affiliation	Links ethnic identities to political leanings or parties.
	Ethnic Generalization	Generalizing ethnic groups with specific traits or behaviors.
	Ethnic Offensive Language	Use of derogatory language targeting specific ethnic groups.
	Threat to Population Narratives	Framing ethnic groups as threats to the national population.
	Ethnic Sexist Narratives	Combining ethnic stereotypes with sexist views.
Super Narrative	Narrative	Description
Migration-related	Migrants Societal Threat	Depicts migrants as a threat to societal norms and security.
Distrust in Institutions	Failed State	Describes a state perceived as failing its citizens and respon bilities.
	Criticism of National Policies	Critique of current governmental policies deemed ineffective harmful.
Distrust in Democratic System	Elections are Rigged	Claims that electoral processes are fraudulent or manipulate
	Anti-Political System	General distrust or rejection of the political system as a who
	Anti-Media	Narratives accusing the media of bias, falsehoods, and manip lation.
	Immigrants Right to Vote	Debates over the voting rights of immigrants in national el tions.
Geopolitics	Pro-Russia	Narratives supportive of Russian policies or perspectives.
	Foreign Interference	Concerns about foreign nations interfering in domestic politi
	Anti-International Institutions	Opposition to international bodies such as the UN or the EU
Anti-Elites	Bilderberg Elites	Focuses on secretive gatherings of elites, suggesting hidd global influence.
	Soros	Centrally features George Soros as a manipulator in glob politics.
	World Economic Forum/Great Reset	Criticism of WEF initiatives perceived as elitist or controllin
	Anti-Semitism	Narratives that inherently carry anti-Semitic sentiments.
	Great Replacement	The theory that elites are replacing native populations wi immigrants.
	Green Agenda	Scepticism or opposition to environmental efforts led by p ceived elites.

Label	Precision	Recall	F1-Score	Support
-1	0.00	0.00	0.00	2
0	0.70	0.69	0.69	119
1	0.00	0.00	0.00	7
2	0.00	0.00	0.00	2
3	0.12	0.14	0.13	7
4	0.37	0.35	0.36	31
5	0.00	0.00	0.00	1
6	0.00	0.00	0.00	1
7	1.00	0.14	0.25	7
9	0.42	0.55	0.48	62
10	0.00	0.00	0.00	1
accuracy			0.54	240
macro avg	0.24	0.17	0.17	240
weighted avg	0.54	0.54	0.53	240

Table 1. BERT Classification Report

Category	Precision	Recall	F1-Score	Support
Anti-EU	0.00	0.00	0.00	7.0
Anti-Elites	0.00	0.00	0.00	2.0
Distrust in democratic system	0.00	0.00	0.00	8.0
Distrust in institutions	0.00	0.00	0.00	36.0
Ethnicity-related	0.00	0.00	0.00	2.0
Gender-related	0.0	0.0	0.0	4.0
Geopolitics	0.0	0.0	0.0	4.0
Migration-related	0.0	0.0	0.0	2.0
None	0.43037974683544300	1.0	0.6017699115044250	102.0
Political hate and polarisation	0.0	0.0	0.0	72.0
Religion-related	0.0	0.0	0.0	1.0
accuracy	0.425	0.425	0.425	0.425
macro avg	0.039125431530494800	0.09090909090909090	0.054706355591311300	240.0
weighted avg	0.1829113924050630	0.425	0.25575221238938100	240.0

Table 2. Binary Chain Classification Report