Analysing and Modelling of Election Narratives

By

Team Charlie

Gaurav Dhande, Munesh Kumar, Nivedita Menon,
Shweta Kakade, Swapnanil Ghosh, Sharath Devanand

Supervisors

Dr. Carolina Scarton and Ibrahim Farha

Module

COM6911: Team Project

University of Sheffield

Faculty of Engineering

Department of Computer Science

Contents

List of Tables

List of Figures

Chapter 1

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.

Employing meticulous data pre-processing techniques, including tokenisation and manual annotation, we meticulously construct a robust training dataset. We then embark on a systematic exploration of various pre-processing 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 to 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.

Chapter 2

Introduction

Social media platforms have become the most important venues for the spread of information in today's digital landscape, frequently blurring the boundaries between reality and fiction. With a particular focus on data from the UK General Elections of 2019[3], this research focuses on creating a sophisticated classifier ready to steer the complexity of digital narratives. The 2019 Elections served as both a powerful political event and a useful examination of digital communication strategies, highlighting the important role that social media plays in influencing public opinion. There is a growing need for robust, automated tools to monitor and capture these narratives as misinformation and polarisation flourish. This project discusses this demand by assembling a model competent in specifying both overarching super-narratives and more fine thematic narratives within tweets, fitted with a novel confidence scoring mechanism.

Throughout the project, annotation played a critical role in training and validating the classification models. A detailed procedure involving double annotations was utilised to assure consistency and depth in comprehending the narratives existing in the dataset. This dual-layer annotation not only delivered a robust basis for the initial training of the models but also permitted for a necessary evaluation of the model's performance against a manually verified standard. So the double annotated set was used as the test set in our classification task.

The preliminary research question of this analysis questions: How effectively can a classifier be prepared to recognise and classify intricate disinformation narratives within social media content during crucial political events? The conclusions of this study demonstrate varying grades of classification triumph, with the BERT model exhibiting moderate precision and recall in specific categories, while other models, including binary chain classifiers, displayed constraints in handling class imbalance and sophistication in narrative classification.

By incorporating advanced computational approaches and a systematic annotation framework, this study contributes a worthwhile understanding of the capabilities and limitations of current machine learning approaches to detect and classify disinformation narratives. The outcomes underscore the complexities of narrative analysis in the digital age and suggest pathways for future enhancements in both methodology and model development.

Chapter 3

Literature Review

3.1 Analysis of COVID-19 Mis/Disinformation Narratives

Kotseva[4] analyzed 58,625 articles from 460 unverified sources on a span of three years to specify and categorise COVID-19 mis/disinformation narratives. This enormous dataset supplied an extensive picture of misinformation movements, echoing the global hierarchy and sophistication of the pandemic's information landscape. Using NLP approaches, the experimenters created a hierarchical codebook comprising 12 super narratives, 51 narratives, and 44 subnarratives, demonstrating prominent themes such as fearmongering, criticism of institutions, and conspiracy theories. Their BERT-based model allowed real-time pursuit of how these narratives developed and spread, delivering worthwhile perspicuity into public perception and behaviour during a public health crisis. Articles were accumulated through the Europe Media Monitor (EMM) system and translated into English, while text clustering was performed using Term Frequency-Inverse Document Frequency (TF-IDF) and Latent Semantic Analysis (LSA). Two annotators manually organised narratives, and the BERT-based classifier was trained on 30,000 annotated articles. The findings revealed that fearmongering overwhelmed early pandemic discourse but slowly subsided after spring 2020 as the understanding of the virus grew. Criticism of authorities like the EU, WHO, and national governments peaked in mid-2020 and crossed with geopolitically set narratives shifting between pro-China, anti-China, pro-Russia, and anti-Russia compositions. Conspiracy theories stayed invariant throughout the pandemic, with periodic spikes, while vaccine-related misinformation surged as vaccines became available, targeting vaccine hesitancy and obligatory vaccination. Despite the exhaustive analysis, limitations included a geographic predilection focusing on Western sources, clustering challenges using TF-IDF algorithms, and translation artefacts guiding to linguistic misclassification. Kotseva's hierarchical codebook shows a structured understanding of recurring mis/disinformation themes, useful for pinpointing patterns and directing classification and counter-narrative strategies.

3.2 Challenges and Frameworks in Narrative Extraction

Santana[10] 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 emphasises 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 tokenisation, normalisation, 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 demand temporal reasoning and entity relation extraction to ensure a 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 categorising and linking narrative elements, directly informing methodologies for detecting and classifying disinformation.

3.3 Designing a System for Formal Narrative Extraction

Metilli, et. al.[5] designed a system for extracting formal narratives from text, emphasising 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 in 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 does not 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 formalising disinformation narratives, and the web-based visualisation interface could inspire similar tools for identifying and analysing mis/disinformation.

3.4 Analysis of Election-Related Disinformation: Insights from a Pan-European Study

Panizio[2] 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 categorised and colour-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 organisations may miss less-publicised 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.

3.5 The Role of Advertising in Online Misinformation

Ahmad[1] examines how online misinformation is financially backed through advertising, with a precise emphasis on the roles of advertisers and digital ad platforms. Descriptive analysis and survey-based investigation studied how misinformation websites secure advertising revenue and how consumer conduct changes when made mindful of this association. The analysis furthermore investigated how decision-makers reacted to learning that their companies promoted misinformation Data was collected from NewsGuard, the Global Disinformation Index (GDI), and Oracle's Moat Pro platform to analyse advertising conduct 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 discovered that 44% of advertisers from a dataset of 42,595 emerged on misinformation sites, and those utilizing digital ad platforms were 10 times more likely to have their advertisements appear on misinformation outlets. When consumers learned about companies advertising on misinformation sites, multiple changed 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 emphasises how advertising platforms amplify disinformation by providing monetization channels, underscoring the importance of understanding these dynamics to refine narrative detection strategies.

3.6 Data-Driven Analysis of Russian Disinformation Narratives in the U.S. Media Environment

The paper by Oates, et. al.[8] offers a data-driven method to catch and scrutinize Russian disinformation narratives and outline their penetration into the U.S. media environment. The authors concentrated on narratives linked to indictments of Nazism against the Ukrainian Azov Battalion and false flag propaganda. By utilising the VAST-OSINT system, they provided an understanding of how Russian disinformation "supply chains" reproduce propaganda across various media platforms. The

VAST-OSINT system compiled and organized suitable online content from over 3 billion URLs, while NLP and network analysis were employed to pinpoint linguistic patterns uniform with Russian propaganda narratives. Case studies concentrated on spreading the "Ukrainian Nazis" narrative associated with the Azov Battalion and claims of U.S. false flag operations. The analysis specified 93 unique URLs promoting the narrative that the Ukrainian Azov Battalion is a neo-Nazi group, especially on far-right U.S. conspiracy websites. The study also discovered examples where Russian propaganda claimed the U.S. was planning false flag operations to justify military action. The authors found out that right-wing websites amplified Russian narratives, while mainstream U.S. media displayed minimal engagement, though niche conspiracy platforms supplied a substantial echo chamber. The analysis concentrated on the English language and right-wing U.S. media, probably neglecting narratives in different languages or political topics, while NLP challenges made identifying precise sources challenging due to deliberate obfuscation. Despite these constraints, the method used and results align presently with the project's goals, displaying how strategic disinformation supply chains can be mapped to acquire vital insights. Understanding which narratives reverberate in different environments is essential for detecting and countering strategic disinformation messaging.

3.7 RESIST 2 Counter-Disinformation Toolkit: A Framework for Recognizing and Fighting Disinformation

The "RESIST 2 Counter-Disinformation Toolkit", prepared by the UK government [9], 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 categorising disinformation, establishing early warning systems, and formulating strategic communications. The RESIST 2 framework stands for Recognise, Early Warning, Situational Insight, Impact Analysis, Strategic Communication, and Tracking Effectiveness. Disinformation is typed into misinformation (without intent to mislead), misuse of valid information, and disinformation (with intent). Case studies from around the globe exhibit the underlying principles in action, emphasising early forewarning via monitoring tools to notice disinformation threats promptly. The framework enables narrative recognition, accentuating the importance of identifying regular themes and symbolism across movements. Strategic communications propose proactive and reactive strategies like pre-bunking and counter-narratives, while impact assessment supplies techniques for analysing disinformation's impacts on policy, stature, and trust. The toolkit's effectiveness depends on consistent implementation across organisations, 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 emphasising proactive communication strategies, like pre-bunking, to align with the project's goals.

3.8 Analysis of International Disinformation During the COVID-19 Pandemic

Damian Milewski[6] analyses the international reach of disinformation during the COVID-19 pandemic, concentrating particularly on China, Russia, and the USA. The author discusses how state or state-funded organisations exploit existing misinformation to formulate narratives that suit their strategic pursuits. Data collection and analysis are established on press and government knowledge using an open-source method, emphasising research, synthesis, and deduction. Case studies spotlight explicit narratives and campaigns from China, Russia, and the USA, emphasising the forms, methods, and tools used. The pandemic constructed immaculate prerequisites for information warfare due to across-the-board public fear and chaos. Disinformation campaigns manipulated existing units and leveraged trusted "super-spreaders." China's narrative emphasised that COVID-19 was initiated outside China while placing the nation as a global leader in pandemic control, using the "three wars" concept to control perception through psychological, media, and legal warfare. Russia's narrative induced fear and distrust through conspiracy theories around Western countries and the USA, while the "Gerasimov Doctrine" underpins Russia's mixed warfare tactics, obscuring the lines between peace and war. The USA aimed mostly to counter Chinese and Russian disinformation, but its messaging was erratic and was missing a comprehensive strategy. Open-source dependency seemingly caps comprehensive wisdom into clandestine disinformation actions, while recognizing sources stays challenging due to intentional obfuscation. Nonetheless, Milewski's paper furnishes a thorough learning of global disinformation drives, crucial for the project's goals. Analysing geopolitical actors' use of disinformation helps inform the project's design for detecting, countering, and enlightening against these narratives.

Chapter 4

Aims & Objectives

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 indicates 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 pre-processing 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 classification accuracy for narratives through the exploration of various pre-processing techniques, such as named entity recognition (NER), aimed at identifying fabricated stories, manipulated statistics, and other specific types of 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 annotation on the tweet dataset to create labelled training data for the model, facilitating supervised learning and improving classification performance.

4. **Objective 4**: Conduct data analysis on human-annotated tweets to extract useful insights into the characteristics and distribution of narratives, informing further model refinement and optimisation strategies.

5. **Objective 5**: 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.

Chapter 5

Methodology

The methodology section details the comprehensive procedures and analytical approaches utilised to model narratives and super narratives guided by tweets from X (formerly Twitter) during the United Kingdom's General Elections in November and December 2019. Since narratives are a core concept in this study, it is essential to establish and introduce a formal definition of the term. "*Narratives*" in the context of the report can be described as linguistic structures that provide a perspective of an individual with respect to a set of events or consequences. This research aims to decipher broad conversation themes and their respective sub-themes from narratives as articulated through public tweets. The sections below provide an account of the data acquisition, structuring, annotation process, and analytical techniques applied.

5.1 Data Acquisition

The quality of classification tasks related to natural language processing is directly associated with the preparation and annotation used to train NLP algorithms. Data for this study was acquired via X's API, by tracking accounts of UK Members of Parliament and targeting tweets from the period of November to December 2019.

5.1.1 Silver Annotations

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 classification model, particularly leveraging Large Language Models (LLMs) to automate the categorisation of narratives and super narratives within tweets.

However, the silver annotation approach encountered significant challenges. Despite the potential for efficiency gains, this method proved ineffective for several reasons. Firstly, the computational demands of running LLMs were high; the infrastructure required to process data continuously and reliably was substantial. Secondly, the output from these models was often inconsistent, with issues such as multiple labels being assigned to a single tweet, which complicated the categorisation 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. This together with the selection of sub-optimal tweets with inadequate context words or tweets containing general statements further complicated the task.

5.1.2 Tweet Selection Criterion

Due to the aforementioned complications, a strategic pivot was necessary. The study shifted towards establishing specific inclusion and exclusion criteria that emphasised the necessity for tweets to be standalone and convey complete thoughts. This change ensured that each tweet could be understood and analysed 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”, and ”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, utilising 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.

5.2 Data Structure

The structured dataset utilised in this study comprises individual tweets sourced from X’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 prioritises 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 X’s official Data Dictionary Standard.

5.3 Data Annotation

5.3.1 Annotators Background

Following the Data Acquisition, several data annotation sessions were scheduled by the authors of this report, all of whom are Master's candidates in Data Analytics at the University of Sheffield. The annotators possess working-level proficiency in English and have backgrounds from various domains within Engineering and are well-equipped to perform a comprehensive annotation.

5.3.2 Annotation Guidelines

To facilitate a nuanced annotation, a hierarchical taxonomy was developed, outlined in a comprehensive codebook accompanying the dataset. The code book served 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 code book, 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.

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 miss-classification or errors and ensuring the accuracy and reliability of the annotated dataset.

To supplement the ease of annotation the annotators were introduced to the GATE Teamware platform that sequentially displayed each tweet with the various super narratives, narratives and confidence scores. Additionally, the platform also had provisions for providing comments should the annotator feel compelled to explicitly mention their reasoning or any additional information.

5.3.3 Annotation Process

A total of 1200 tweets were annotated by all the researchers across multiple annotating sessions. Throughout the annotation process, each of the 1200 tweets in the dataset underwent individual evaluation by the annotators, each responsible for labelling 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 180 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 labelled dataset.

5.3.4 Inter-annotator Disagreements

When multiple annotators were involved in annotating the test dataset, naturally there were disagreements regarding tweet labels. Such disagreements were discussed among the annotators each putting forth their perspective on how they interpreted the tweet. If a justification provided by one of the annotators resonated with both of them it was selected as the label. However, if none of them could still agree, a third intermediary (another annotator) was called upon to resolve and would act as the deciding authority over the eventual label for the tweet.

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

5.4 Data Analysis

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

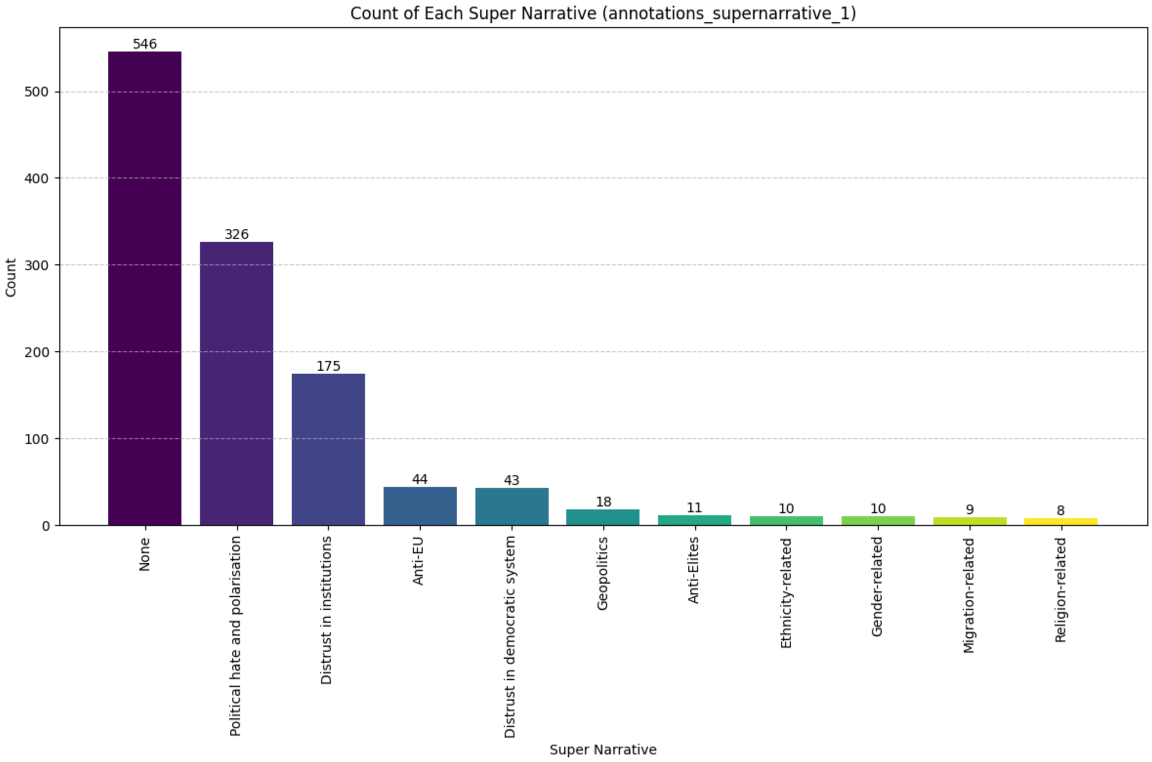


Figure 5.1: Count of Supernarratives

The analysis revealed a significant dominance of tweets classified under ”No narrative provided”, accounting for 546 instances. This category includes tweets that lacked a discernible narrative or were strictly informational, suggesting that a substantial portion of the discourse was centred 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 disengage-

Super Narratives	**Frequency**
None	546
Political Hate and Polarization	326
Distrust in Institutions	175
Anti-Eu	44
Distrust in Democratic System	43
Geopolitics	18
Anti-Elites	11
Ethnicity-Related	10
Gender-Related	10
Migration-Related	9
Religion-Related	8

Table 5.1: Distribution of Narratives across Tweets

ment 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 scepticism 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 behaviour and election outcomes.

An exhaustive table that shows the distribution of tweets across different super-narratives across is shown in Table 5.1 and a graph for the same can be seen in Figure 5.1.

5.5 Model Selection

5.5.1 Embeddings

BERT is a pre-trained language model developed by Google. It’s bidirectional, meaning it considers context from both the left and right sides 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 Meta AI. It’s trained on diverse datasets and tasks simultaneously, allowing it to generalise 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 labelling, 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.

5.5.2 Traditional 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. Table 5.2 highlights the performance with metrics - accuracy and F1 of the models that were implemented.

Model	**Accuracy**	**F1-Score**
Multinomial NB	0.29	0.21
Logistic	0.27	0.22
Random Forest	0.31	0.25
Linear SVM	0.24	0.22
MultiLayer Perceptron	0.25	0.23
Gradient Boost	0.25	0.24
Adaboost	0.3	0.15
KNN	0.28	0.25
Decision Tree	0.25	0.24
Bagging	0.25	0.21
XG Boost	0.24	0.22

Table 5.2: Metrics of Models

5.5.3 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 utilised 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 labelled 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.

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.

Chapter 6

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 contrast, the binary chain classifier displays uniformly poor performance across all categories (Figure 6.1), 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.

Class	**Precision**	**Recall**	**F1-score**	**Support**
-1	0.00	0.00	0.00	3
0	0.52	0.70	0.59	67
1	0.00	0.00	0.00	11
2	0.10	0.33	0.15	3
3	0.00	0.00	0.00	9
4	0.21	0.35	0.26	31
5	0.00	0.00	0.00	2
7	0.00	0.00	0.00	5
8	0.00	0.00	0.00	2
9	0.35	0.20	0.25	46
10	0.00	0.00	0.00	1
Total				**180**
Accuracy	**0.38**			
Macro avg	0.11	0.14	0.11	
Weighted avg	0.32	0.38	0.33	

Table 6.1: BERT Classification Report

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.

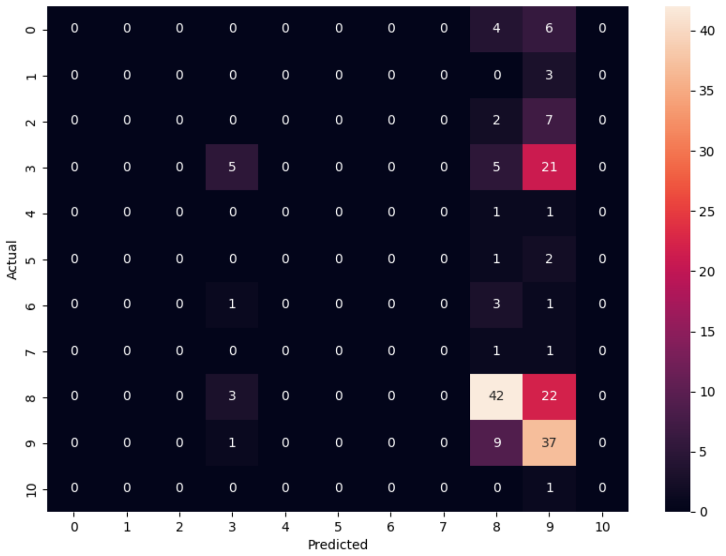


Figure 6.1: Confusion Matrix for Binary Chaining Classifier

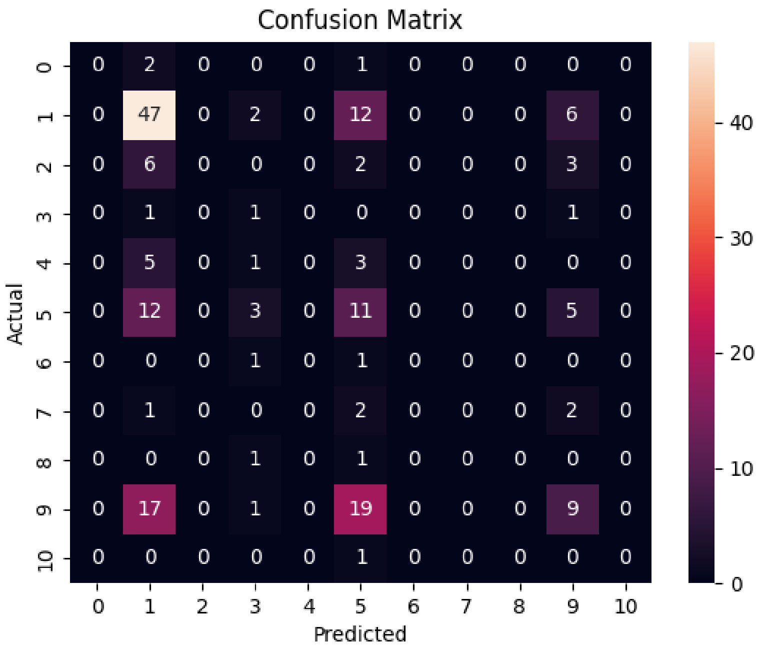
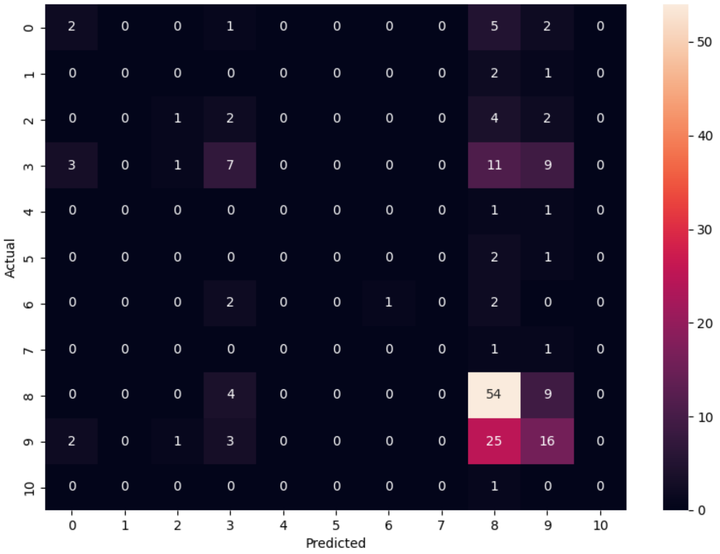
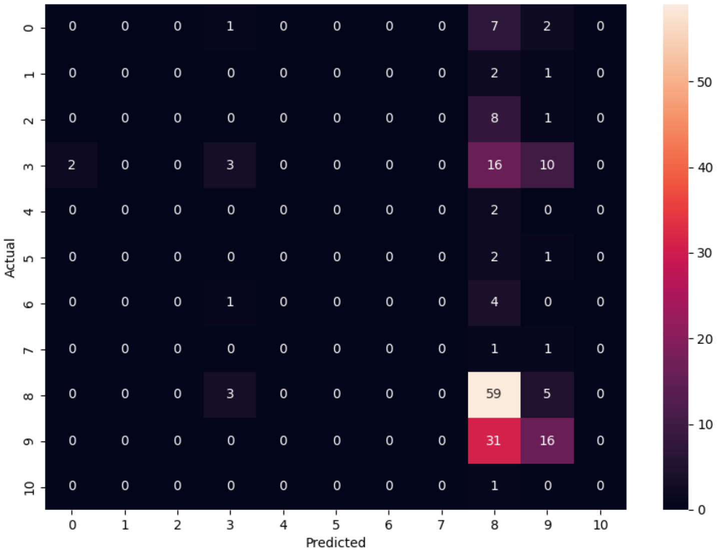


Figure 6.2: Confusion Matrices for BERT



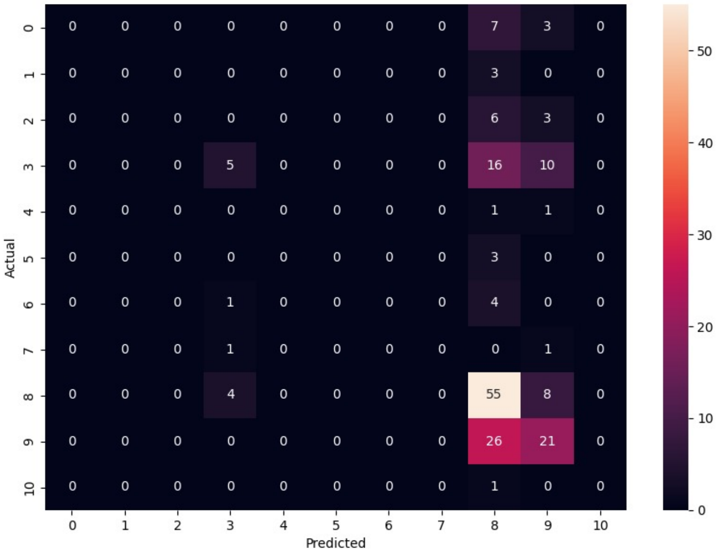
	0	1	2	3	4	5	6	7	8	9	10
0	2	0	0	1	0	0	0	0	5	2	0
1	0	0	0	0	0	0	0	0	2	1	0
2	0	0	1	2	0	0	0	0	4	2	0
3	3	0	1	7	0	0	0	0	11	9	0
4	0	0	0	0	0	0	0	0	1	1	0
5	0	0	0	0	0	0	0	0	2	1	0
6	0	0	0	2	0	0	1	0	2	0	0
7	0	0	0	0	0	0	0	0	1	1	0
8	0	0	0	4	0	0	0	0	54	9	0
9	2	0	1	3	0	0	0	0	25	16	0
10	0	0	0	0	0	0	0	0	1	0	0

Figure 6.3: Confusion Matrices for MLP



	0	1	2	3	4	5	6	7	8	9	10
0	0	0	0	1	0	0	0	0	7	2	0
1	0	0	0	0	0	0	0	0	2	1	0
2	0	0	0	0	0	0	0	0	8	1	0
3	2	0	0	3	0	0	0	0	16	10	0
4	0	0	0	0	0	0	0	0	2	0	0
5	0	0	0	0	0	0	0	0	2	1	0
6	0	0	0	1	0	0	0	0	4	0	0
7	0	0	0	0	0	0	0	0	1	1	0
8	0	0	0	3	0	0	0	0	59	5	0
9	0	0	0	0	0	0	0	0	31	16	0
10	0	0	0	0	0	0	0	0	1	0	0

Figure 6.4: Confusion Matrices for SVM



Actual \ Predicted	0	1	2	3	4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	0	7	3	0
1	0	0	0	0	0	0	0	0	3	0	0
2	0	0	0	0	0	0	0	0	6	3	0
3	0	0	0	5	0	0	0	0	16	10	0
4	0	0	0	0	0	0	0	0	1	1	0
5	0	0	0	0	0	0	0	0	3	0	0
6	0	0	0	1	0	0	0	0	4	0	0
7	0	0	0	1	0	0	0	0	0	1	0
8	0	0	0	4	0	0	0	0	55	8	0
9	0	0	0	0	0	0	0	0	26	21	0
10	0	0	0	0	0	0	0	0	1	0	0

Figure 6.5: Confusion Matrices for Stacking Classifier

In Figures 6.2, 6.3, 6.4 and 6.5, you can see the confusion matrices for all the classifiers that we used like, MLP, BERT, SVM and Stacking. The SVM model appears to perform slightly better overall, with a slightly higher accuracy (53.33 %) compared to the MLP (49.17%) and Stacking (48.75%) models. This suggests that for this specific dataset, the SVM might be handling the feature space more effectively. The None category (likely representing no specific narrative or sentiment) has notably high recall in all models, especially in SVM (89.22%), indicating that the SVM model is particularly good at identifying the dominant class but might be at the cost of other smaller classes.

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Anti-EU	0.00	0.00	0.00	10
Anti-Elites	0.00	0.00	0.00	3
Distrust in democratic system	0.00	0.00	0.00	9
Distrust in institutions	0.50	0.16	0.24	31
Ethnicity-related	0.00	0.00	0.00	2
Gender-related	0.0	0.0	0.0	3
Geopolitics	0.0	0.0	0.0	5
Migration-related	0.0	0.0	0.0	2
None	0.62	0.63	0.62	67
Political hate and polarisation	0.36	0.79	0.50	47
Religion-related	0.0	0.0	0.0	1
Total				**180**
Accuracy	**0.47**			
Macro Average	0.13	0.14	0.12	
Weighted Average	0.41	0.47	0.40	

Table 6.2: Evaluation Metrics for Binary Chain Classification

Class	**Precision**	**Recall**	**F1-score**	**Support**
Anti-EU	0.00	0.00	0.00	10
Anti-Elites	0.00	0.00	0.00	3
Distrust in democratic system	0.00	0.00	0.00	9
Distrust in institutions	0.38	0.10	0.15	31
Ethnicity-related	0.00	0.00	0.00	2
Gender-related	0.00	0.00	0.00	3
Geopolitics	0.00	0.00	0.00	5
Migration-related	0.00	0.00	0.00	2
None	0.44	0.88	0.59	67
Political hate and polarisation	0.43	0.34	0.38	47
Religion-related	0.00	0.00	0.00	1
Total				**180**
Accuracy	**0**.43			
Macro avg	0.11	0.12	0.10	
Weighted avg	0.34	0.43	0.35	

Table 6.3: SVM Evaluation Metrics

Class	**Precision**	**Recall**	**F1-score**	**Support**
Anti-EU	0.29	0.20	0.24	10
Anti-Elites	0.00	0.00	0.00	3
Distrust in democratic system	0.33	0.11	0.17	9
Distrust in institutions	0.37	0.23	0.28	31
Ethnicity-related	0.00	0.00	0.00	2
Gender-related	0.00	0.00	0.00	3
Geopolitics	1.00	0.20	0.33	5
Migration-related	0.00	0.00	0.00	2
None	0.50	0.81	0.62	67
Political hate and polarisation	0.38	0.34	0.36	47
Religion-related	0.00	0.00	0.00	1
Total				**180**
Accuracy	**0.45**			
Macro avg	0.26	0.17	0.18	
Weighted avg	0.41	0.45	0.40	

Table 6.4: MLP Evaluation Metrics

Class	**Precision**	**Recall**	**F1-score**	**Support**
Anti-EU	0.00	0.00	0.00	10
Anti-Elites	0.00	0.00	0.00	3
Distrust in democratic system	0.00	0.00	0.00	9
Distrust in institutions	0.45	0.16	0.24	31
Ethnicity-related	0.00	0.00	0.00	2
Gender-related	0.00	0.00	0.00	3
Geopolitics	0.00	0.00	0.00	5
Migration-related	0.00	0.00	0.00	2
None	0.45	0.82	0.58	67
Political hate and polarisation	0.45	0.45	0.45	47
Religion-related	0.00	0.00	0.00	1
Total				**180**
Accuracy	**0.45**			
Macro avg	0.12	0.13	0.12	
Weighted avg	0.36	0.45	0.37	

Table 6.5: Evaluation Metrics for the Stacking Classifier

Chapter 7

Discussion and Conclusion

We had chosen the Large Language Model approach over the model clustering for silver annotations for increased accuracy and to ensure all the super narratives are taken into account. LLaMA was considered for the aforementioned task where a set of binary questions for each narrative was posed and the results were retrieved. The provided tweets were in a JSON format which had a high computational time to process in a Python script file, this shifted us to work with the CSV format of the provided data which presented with a low parsing time. The tweets were initially equally divided among the team members to annotate with LLaMA2[7]. Careful consideration of the prompt was used to obtain only the binary value as a result. Another significant hurdle was overcoming the profanity censor built into LLaMA2. This resulted in switching to the uncensored version of the LLaMA2 model. As the allotted tweets for our group comprised majorly of retweets, the silver annotations of the other team were later considered for data annotation.

Double annotations were done by teaming up with another team member of the group and annotation the tweets the second time. The double annotations brought immense understanding to the tweet and presented all of us with a common perspective in the annotation. These double annotations were later used as a test set for comparing the model performances for classification.

Double annotations added a new depth to each narrative and super narrative and how each of them was conceived. The quantitative metric of confidence level aided in laying a formal definition for improved interpretation of the annotations. The option of secondary narrative and super narrative brought about a new dimension into how the tweets are classified and let to an understanding that a tweet is not restricted to a single narrative-super narrative label.

Each team member was given 200 tweets that were to be annotated and used for the training set for future models which was spread across 3-5 weeks. A lot of discussion was based on how the tweets would affect the model design and the impact of the secondary narratives on the model. During manual annotations, the balance across the narrative label was focused and it could be inferred that 4 out of the 10 narratives were majorly considered throughout the annotation phase. The 'None' option was considered to be one of the most selected options as there were tweets related to campaigning and other non-election-related tweets that would not fall under any of the narratives.

Challenges in Minority Classes:
Classes with fewer samples, such as 'Anti-Elites', 'Ethnicity-related', 'Gender-related', 'Geopolitics', 'Migration-related', and 'Religion-related', have very low to zero precision and recall across all models. This suggests an issue with class imbalance where models are failing to adequately learn features specific to less represented classes. Techniques like oversampling the minority class, adjusting class weights, or using anomaly detection approaches might help improve performance on these classes.

Comparison and Strategy for Improvement:
The precision, recall, and F1 scores for categories like ‘Political hate and polarisation’ and ‘Distrust in institutions’ vary significantly across models, indicating that different models capture different aspects of the data. For example, MLP seems to slightly underperform compared to SVM in precision for the ’Political hate and polarisation’ category but has similar recall.
An ensemble approach, as attempted with the Stacking classifier, did not significantly outperform the individual models, which might indicate that the combination of these particular models does not complement each other effectively or the meta-model needs fine-tuning.
Experimenting with different combinations of base classifiers or more sophisticated meta-classifiers in the stacking approach could potentially yield better results.
In conclusion, while the SVM classifier provided the best overall metrics among the models, significant room for improvement remains, particularly in effectively classifying minority classes and enhancing the generalization capabilities of the models. Further experimentation with data preprocessing, model selection, and hyperparameter tuning is recommended to address these challenges.

Annotation Quality and Consistency:
The annotations might vary significantly in quality and consistency depending on the annotator, which could explain some of the discrepancies and lower performances for certain classes in the model predictions. Inconsistent or subjective annotations can lead to a model learning incorrect patterns, thus affecting its ability to generalize well across unseen data. This might be particularly noticeable in classes with low support, where a few mislabeled examples can significantly skew the results. Implement a review or adjudication process where annotations are checked and validated by multiple annotators or an expert, especially for underperforming classes, to enhance annotation reliability and model learning.

Annotation Distribution and Class Imbalance:
Some classes, such as ‘Anti-EU’, ‘Distrust in democratic system’, and ‘Religion-related’, have very few examples (low support), leading to zero performance in metrics, indicating a class imbalance in the annotated dataset. The imbalance leads to insufficient learning opportunities for the model regarding rare classes, causing the model to be biased towards more frequent classes. Augment the dataset with more examples of underrepresented classes either by gathering more data or through techniques like text augmentation. Alternatively, adjust the training process to give more weight to these minority classes.

Complexity of Textual Features in Tweets:
Given the varied performance across different categories, it is likely that some categories of annotations are associated with more straightforward linguistic patterns than others. For example, classes like ‘None’ and ‘Political hate and polarisation’ are easier for models to predict, possibly due to more distinct keywords or phrases. The complexity and subtlety of language used in tweets related to more nuanced categories like ‘Distrust in institutions’ or ‘Geopolitics’ might not be captured well by the current feature extraction method.

Chapter 8

Future Scope

In this project, we considered the 2019 elections for the classification of narratives. As observed in that set of narratives, a subset of it revolved around the major narrative of ‘Brexit’. The manifesto and the agenda of all the political parties revolved around their stand on the Brexit issue in the year 2019. The future scope of this project would be to identify the one-off issue that differentiates all the parties and potentially classify it appropriately. This means that we would ideally want to generalize over all the elections and retain the recurring narratives such as political hate and polarization and modify narratives that are constantly updating over the years.

The second agenda for the future scope is to incorporate Large Language Models such as LLaMA3 and RoBERTa in the model classification as they are built with large datasets. Minor tweaking of these LLMs would lead to an efficient narrative classifier that supports the higher objectives of the VIGILANT project.

Finally, the code-book proved to be ineffective in terms of certain narratives and super-narratives not being used throughout the annotation phase. The next step would be to filter out unused narratives and add new narratives and super-narratives that are more commonly emerging. This would lead to a more uniform label distribution across the tweets.

There was also a clear sign of imbalance in the dataset which should be taken into account during further processing steps. A balance should be initiated either through oversample, undersampling, or other balancing operations to ensure a distributed dataset. The required uniformity in the distribution would lead to improved model performance for understanding all the labels and tweaking the weights accordingly.

Bibliography

[1] Wajeeha Ahmad et al. *The role of advertisers and platforms in monetizing misinformation: Descriptive and experimental evidence.* Tech. rep. National Bureau of Economic Research, 2024.

[2] *Disinformation Narratives During the 2023 Elections in Europe.* URL: `https://edmo.eu/wp-content/uploads/2023/10/Narratives-Report.pdf`.

[3] Genevieve Gorrell et al. "Which politicians receive abuse? Four factors illuminated in the UK general election 2019". In: *EPJ Data Science* 9.1 (2020), p. 18.

[4] Bonka Kotseva et al. "Trend analysis of COVID-19 mis/disinformation narratives–A 3-year study". In: *Plos one* 18.11 (2023), e0291423.

[5] Daniele Metilli, Valentina Bartalesi, Carlo Meghini, et al. "Steps Towards a System to Extract Formal Narratives from Text." In: *Text2Story@ ECIR*. 2019, pp. 53–61.

[6] Damian Milewski. "The analysis of narratives and disinformation in the global information environment amid Covid-19 pandemic". In: (2020).

[7] Thanh Thi Nguyen, Campbell Wilson, and Janis Dalins. "Fine-tuning llama 2 large language models for detecting online sexual predatory chats and abusive texts". In: *arXiv preprint arXiv:2308.14683* (2023).

[8] Sarah Oates, Doowan Lee, and David Knickerbocker. "Data Analysis of Russian Disinformation Supply Chains: Finding Propaganda in the US Media Ecosystem in Real Time". In: *Oates, Sarah, Doowan Lee, and David Knickerbocker* (2022).

[9] *RESIST 2 Counter-Disinformation Toolkit.* URL: `https://gcs.civilservice.gov.uk/publications/resist-2-counter-disinformation-toolkit/`.

[10] Brenda Santana et al. "A survey on narrative extraction from textual data". In: *Artificial Intelligence Review* 56.8 (2023), pp. 8393–8435.

Appendix A

Appendix

Section A: Applicant Details

Name: Nivedita Menon
Email Address: nmenon2@sheffield.ac.uk
Research Project Title: Analysing Disinformation Narratives during Elections
Has your research project undergone academic review in accordance with the appropriate process? YES

Section B: Basic Information

1. Application Details

Co-applicant(s):

- Name: Gaurav B Dhande, Email: gbdhande1@sheffield.ac.uk
- Name: Munesh Kumar, Email: mkumar3@sheffield.ac.uk
- Name: Sharath Devanand, Email: sdevanand1@sheffield.ac.uk
- Name: Shweta Kakade, Email: skakade1@sheffield.ac.uk
- Name: Swapnanil Ghosh, Email: sghosh6@sheffield.ac.uk

2. Proposed Project Duration:
Start Date (of data collection): 18/03/2024
Anticipated End Date (of project): 18/05/2024
3. Project Code (where applicable): [To be filled]
Project Externally Funded? No

4. Suitability

Question	Answer
Takes place outside UK?	No
Involves NHS?	No
Health and/or social care human-interventional study?	No
ESRC funded?	No
Likely to lead to publication in a peer-reviewed journal?	Yes
Led by another UK institution?	No
Involves human tissue?	No
Clinical trial or a medical device study?	No
Involves social care services provided by a local authority?	No
Is social care research requiring review via the University Research Ethics Procedure	No
Involves adults who lack the capacity to consent?	No

2. Indicators of Risk

Involves Potentially Vulnerable Participants? No
Involves Potentially Highly Sensitive Topics? Yes

- Race or ethnicity
- Political opinion
- Religious, spiritual or other beliefs
- Criminal or illegal activities
- Political asylum
- Conflict situations
- Personal violence

Section C: Summary of Research

1. Aims and Objectives

The objective of the project is to analyse the narrative from a given text; then classify it into a category of super-narrative and narrative.

2. Methodology

The first step is data collection/acquisition using Twitter/X. Data annotation will follow, focusing on political discourse.

3. Risk to Researchers

Discusses the potential risks to researchers, especially when analyzing sensitive content, and how these are managed.

Section D: About the Participants

How will you identify the potential participants? Participants are the members of the project team from the Department of Computer Science.

Section E: About the Data

1. Data Processing

Data will be processed and analyzed ensuring confidentiality and compliance with GDPR.

3. Data Confidentiality

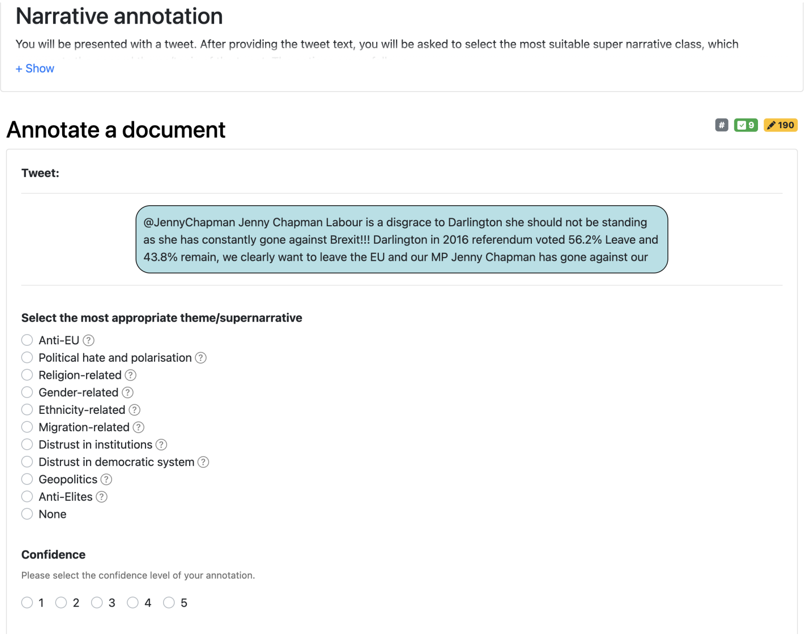
Details on how the confidentiality of participant data will be ensured.

Section F: Supporting Documentation

Information & Consent Forms: Yes

Section G: Data Annotation Tool

Description of the Tool: The following screenshot illustrates Gate Teamware, the tool we used for annotating data. This tool displays tweets alongside a list of super narratives. Our task was to select the most suitable super narrative for each tweet. Once a super narrative was chosen, its respective narratives were shown, and we selected the appropriate narrative. We then rated our confidence in the selections on a scale from 1 (low) to 5 (high). Additionally, we could add comments and justify our selections in the provided boxes before submitting the annotation.



Narrative annotation

You will be presented with a tweet. After providing the tweet text, you will be asked to select the most suitable super narrative class, which

+ Show

Annotate a document

Tweet:

@JennyChapman Jenny Chapman Labour is a disgrace to Darlington she should not be standing as she has constantly gone against Brexit!!! Darlington in 2016 referendum voted 56.2% Leave and 43.8% remain, we clearly want to leave the EU and our MP Jenny Chapman has gone against our

Select the most appropriate theme/supernarrative

○ Anti-EU ?
○ Political hate and polarisation ?
○ Religion-related ?
○ Gender-related ?
○ Ethnicity-related ?
○ Migration-related ?
○ Distrust in institutions ?
○ Distrust in democratic system ?
○ Geopolitics ?
○ Anti-Elites ?
○ None

Confidence

Please select the confidence level of your annotation.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Figure A.1: Screenshot of Gate Teamware