MEDIA AND THE MIGRATION DEBATE: A MACHINE-LEARNING APPROACH TO SENTIMENT ANALYSIS ON TWITTER

MELANIE WATERHAM

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY

DEPARTMENT OF
COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

STUDENT NUMBER

424235

COMMITTEE

Chris Emmery MSc
Henry Brighton PhD

LOCATION

Department of Cognitive Science & Artificial Intelligence
School of Humanities and Digital Sciences
Tilburg University, Tilburg, The Netherlands

DATE

January 13, 2023

WORD COUNT

8572

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor Chris Emmery for his efforts in helping me collect the data, and for his continuous support, availability, and inspiration throughout the thesis period. Furthermore, I wish to thank my friend and fellow student Edoardo Saldarelli, whose pragmatic mindset and advice proved invaluable in my completing this project.

MEDIA AND THE MIGRATION DEBATE: A MACHINE-LEARNING APPROACH TO SENTIMENT ANALYSIS ON TWITTER

MELANIE WATERHAM

Work on this thesis did not involve collecting data from human participants or animals. The original owner of the data and code used in this thesis retains ownership of the data and code during and after the completion of this thesis. The author of the thesis acknowledges that they do not have any legal claim to this data. The data from this thesis is obtained from Twitter, using the Twitter Application Programming Interface (API). The code used in this thesis is publicly available at: https://github.com/melaniewaterham/thesis.

Abstract

This thesis presents a machine-learning approach to analyse migration sentiment in news outlets and their audience on Twitter in the Netherlands and Italy. It aims to answer the research question: *'To what extent can a rule-based model versus a neural network language model identify polarisation in tweets about migration?'*. The results suggest that a pre-trained BERT-based neural network model can accurately predict sentiment in news tweets, with adequate data pre-processing techniques likely improving results on user replies. The model outperforms the baseline setup with the lexicon-based sentiment analysis tool VADER. Findings suggest that not tone but salience of the topic in Twitter media evokes the most intense reactions, with negative sentiment dominating the user replies in both countries.

1 INTRODUCTION

'61% supports asylum policy Netherlands' – proclaimed NOS Teletekst on Twitter in September 2022, quoting a survey (Van der Schelde, 2022) by I&O Research. Dutch public service and news broadcaster NOS had commissioned the survey on public attitudes towards asylum seekers, but was met with overwhelmingly negative responses upon sharing its findings in a tweet. Some users expressed concern about the phrasing of the survey questions, or whether the sample was representative of the Dutch population. Others accused the outlet of spreading lies and propaganda, or used the opportunity to present alternative figures by Ipsos that instead appeared to call for an 'immigration stop'. Two days later, NOS published a follow-up article (NOS Nieuws, 2022) to account for the difference in findings quoted in the original tweet and the yearly Ipsos figures that were referenced in the comments: the seemingly conflicting results could simply be attributed to the way the questions were phrased. 'It is entirely possible for people to think that we generally have an obligation to provide asylum seekers with decent shelter accommodations, while simultaneously observing that as a country, we currently do not manage to do so', states researcher Asher van der Schelde. This interpretation was endorsed by an Ipsos representative: 'Prior research shows considerable support for the general principle of providing shelter for refugees, hence the current asylum crisis may well have affected public sentiment about restricting the influx' (NOS Nieuws, 2022). A different explanation provided by the article is that I&O refers to international treaties on migration in the survey, which Ipsos did not do, stating that 'the context that is provided can influence the responses to a question' (NOS Nieuws, 2022).

1.1 *Background*

Migration has consistently been one of the most divisive social issues around the globe (Dennison & Geddes, 2021). While sometimes considered vital to the development of a country, an influx of migrants is often perceived as a threat to national values, cultural identity, economic stability, and security by its native population (Dennison & Geddes, 2019). According to the Migration Policy Institute, there are five main factors that contribute to anxiety about immigration in a society (Papademetriou & Banulescu-Bogdan, 2016):

1. A rapid pace of social change;
2. The perception of immigrants as competitors for resources, especially in areas of economic hardship;
3. A view of culturally distinct immigrants as a threat to norms and values;
4. Security concerns, or the association between migrants and acts of terrorism and crime;
5. Lack of control, and loss of confidence in governments and elites to control inflows and ensure successful integration of migrants.

A peak in refugee arrivals in Europe brought these topics to the centre of public debate in 2015, as individual countries struggled to deal with the societal, political, and administrative implications of the more than 1.3 million asylum applications registered on the continent in that year (European Asylum Support Office, 2015). Until the armed invasion of Ukraine in February 2022, this statistic reflected the highest number of displaced persons in Europe since World War II. The public perception of these events, now commonly referred to as the 2015 European refugee crisis, has contributed to high-impact political decisions and turning points in some countries affected by them (Dennison & Geddes, 2019; Rowe, Mahony, Graells-Garrido, Rango, & Sievers, 2021). A prominent example is the 2016 Brexit referendum in the United Kingdom that lead to the decision of the country to leave the European Union four years later, a question widely presented as a matter of 'gaining back control' by the Leave campaign. In other European countries, it has driven support for political parties with a nationalist and right-wing populist agenda, often accompanied by an anti-immigration stance. In Italy, the general elections of 2018 saw the defeat of mainstream parties Forza Italia (FI) and the social-democratic PD, with challenger parties M5S and Lega achieving their best electoral results in history on a mixture of progressive economic goals and a conservative stance on immigration. In an analysis of Italian voter attitudes and the Twitter campaign of political parties in the months leading up to the elections, it was found that the level of support for the anti-immigration position was more than 80% in M5S voters, and 90% or more for voters of FI, Lega, and Fratelli d'Italia, with anti-immigrant goals being a crucial point for many centre-right parties (Maggini & Chiaramonte, 2019). In fact, a vast majority of voters of all but two small parties favoured a reduction of refugee inflows in the country (Maggini & Chiaramonte, 2019).

1.2 *Definitions*

In public debate, the terms 'asylum seeker', 'refugee', and 'migrant' are often used interchangeably, but wrongfully so. 'Asylum seekers' are individuals who have submitted a request for asylum outside of their home country, with the outcome still pending. Should the asylum procedure be completed with success, then the individual is granted protection in the country of application. 'Refugee' status is one such kind of protection, but other types of humanitarian protection also exist within the European asylum framework. Nevertheless, not all applications prove successful: during the peak in 2015, 49% of decisions issued at first instance were positive (European Asylum Support Office, 2015). A negative decision does not grant the claimant refugee status, but instead entails the obligation to leave the country – if an individual remains, they become an undocumented migrant. This definition also applies to individuals who choose not to register an asylum claim in the first place, e.g. due to a prolonged waiting period, the intention to only transit through a given country and apply elsewhere, or the awareness that there is little prospect of their claim proving successful (OECD, 2015). Finally, 'migrant' is the generic term for any person moving to another country with the intention to remain there for a defined minimum period of time. This includes permanent migrants, temporary migrants (with a valid residence permit or visa), asylum seekers (with the possibility of becoming humanitarian migrants), and also undocumented migrants. For the sake of consistency, this thesis uses the umbrella term 'migrants' as a way to refer to all aforementioned groups.

1.3 *Migration Policy, Media and Sentiment towards Migrants*

The impact on the national agenda and the complexity of the problem left an opening for the media to shape public understanding of what the 2015 refugee crisis meant for their country (Greussing & Boomgaarden, 2017). In fact, division on this topic is often attributed to media framing. Media and news messages can set the tone for topics of public debate, and by virtue of their use in mass communication, contribute to the perception of these topics. Meltzer et al. (2021) found that the visibility or *salience*, the *framing* of and the *sentiment* in coverage of migration clearly influence public policy preferences. For example, a highly polarised discussion about European labour migration and alignment with partisan identities are generally understood to have contributed to the outcome of the Brexit referendum (Heidenreich, Eberl, Lind, & Boomgaarden, 2020; Meltzer et al., 2021; Rowe et al., 2021). A related factor contributing to polarisation are partisan narratives and misconceptions originating from patterns of information sharing on social media (Menshikova & Van Tubergen, 2022; Rowe et al., 2021). Social media are often believed to comprise a main channel leading to selective exposure to information, thereby reinforcing pre-existing beliefs and affecting the public perception of migration (Rowe et al., 2021).

1.4 *Aims and Relevance*

This thesis intends to explore polarisation on the topic of migrants and migration policy. Specifically, it aims to analyse the differences in migration-related discourse between established media channels and general users on social media. To achieve this goal, two models will be utilised and their performances compared: the lexicon and rule-based sentiment analysis tool VADER (*Valence Aware Dictionary and sEntiment Reasoner*) and the state-of-the-art, neural network-based language model known as RoBERTa. Using these models, we will analyse Twitter threads from various news outlets and their audience to visualise the distribution of migration-related sentiment in different user groups. We will attempt to answer the following research question:

> *To what extent can a rule-based model versus a neural network language model identify polarisation in tweets about migration?*

We define *polarisation* as a high prevalence of strongly positive and negative sentiment on a topic, as opposed to solely neutral statements. Sentiment is measured by its direction or *polarity* (either negative, neutral, or positive), and its *intensity* (ranging from moderate to the most extreme sentiments). Much of the accuracy of our results is dependent on the data collection strategy, together with the performance of our sentiment analysis model. Our models of choice have been developed specifically on social media data and are able to detect features such as emojis (Hutto & Gilbert, 2014) and, to varying degrees, contextual information to accurately determine the valence of a tweet. The current state-of-the-art model in sentiment analysis, RoBERTa, significantly outperforms all other existing methods if trained properly (Liu et al., 2019). A challenge for our task, however, is the fact that the sentiment of a tweet does not reflect the author having a corresponding view on migration. For instance, a tweet with positive valence (e.g. 'Great!') could in fact be a favourable reaction to a more restrictive migration policy, sarcasm, or pleasure derived from the mockery of the original tweet. In the first of three guiding sub-questions that form the backbone of our analysis, we evaluate the performance of both models on a labelled subset of our Twitter data:

> *Q.1: Which performs best on the task of identifying a Twitter user's sentiment about migration; a rule-based classifier utilising the VADER lexicon, or a pre-trained Transformer model?*

Secondly, we investigate the relationship between a country's political direction, media narratives and public opinions on migration. This is achieved by comparing tweets from two European countries with differing political landscapes and dominant narratives about migration. Italy has taken a determined stance on migration in previous years, culminating in the centre-right coalition led by conservative anti-immigration party Fratelli d'Italia (FDI) taking a majority vote in the 2022 general elections, with FDI becoming the largest in parliament with 26% of the vote. By contrast, the Netherlands has been governed by a centre-right coalition for over a decade, with no anti-immigration or far-right representation in the government. We therefore assume polarisation to be greater in our sample of tweets from Italy than from the Netherlands. This hypothesis is addressed by the second guiding question:

> *Q.2: Does the degree of polarisation differ between countries?*

As traditional media convey facts, albeit selected and edited to meet a medium's journalistic criteria (Heidenreich et al., 2020; Theorin & Strömbäck, 2020), it is hypothesised that sentiment in media tweets remains relatively neutral overall, with perhaps slight variation in tone between outlets and tweets. By contrast, commentary by individual users is presumed to be of a more personal and opinionated nature, in particular on a divisive topic such as immigration. Hence, tweets by the public are expected to carry more strongly positive and negative sentiment. Due to the effects of media on migration attitudes (Meltzer et al., 2021; Theorin & Strömbäck, 2020), we also expect readers of more biased outlets to express stronger sentiment in their own tweets. These assumptions are investigated through the third sub-question:

> *Q.3: Is there a difference in the direction and intensity of migration sentiment between various news outlets on Twitter, and how does this compare to their audiences reflecting on the same topic?*

The societal relevance of this project is two-fold. Studying sentiment polarity and intensity within news media content and its consumers can identify areas of heightened polarisation in public discourse. This can set the agenda for measures to combat the emergence of so-called echo chambers on social media, and instead facilitate a more civilised and constructive discussion of complex social issues. Due to the key role of social media in the spread of information and the formation of opinion, this can also be the first step in attempting to normalise the political climate in times of crisis or uncertainty. Such initiatives benefit individuals, but are also likely to contribute to better acceptance of foreign communities and greater social cohesion in countries affected by migration (Dennison & Geddes, 2021; Rowe et al., 2021). Together with prior work on migration sentiment and migrant assimilation, the methods and findings proposed in this analysis can help policymakers gain insight into effective communication strategies, monitor and respond timely to concerns of the public, and make data-driven policy decisions (Menshikova & Van Tubergen, 2022) to improve the successful integration of migrants into society.

The scientific contributions of this thesis lie in the novelty of the approach. Contrary to the study of traditional news sources, the analysis of migration-related news and replies on social media allows for an investigation into valenced news reporting and their effect on readers' emotional state, as demonstrated by the sentiment scores of selected user groups from our sample. Experiments with Twitter data in Italian or Dutch are scarce, and interpreting our findings through a cross-country comparison contributes to the existing body of comparative research on the influence of media on anti-migration sentiment, and its relation to political developments. In addition, the findings will shed light on the suitability of a pre-trained multi-lingual Transformer model for the specific use case of sentiment analysis on migration, and potentially other divisive Twitter topics.

2 RELATED WORK

Early studies on migration-related sentiment (Simon & Lynch, 1999; Zimmermann, Bauer, & Lofstrom, 2000) are generally based on survey data. They are often comparative analyses, with survey responses from different countries presented in the historical and current-day context of their migration policies. Sentiment data is acquired from national but also international opinion polls, such as the Gallup World Poll, World Values Survey, Eurobarometer, and European Community Household Panel. Besides highlighting important migration-related issues according to public opinion, they seek to find explanations for geographical and temporal differences in sentiment. From these sources, it can be concluded that negative attitudes towards immigrants are a global phenomenon, but that these feelings frequently change over time, as well as in their specific makeup. A Canadian study (Wilkes, Guppy, & Farris, 2008) of changes in sentiment over a 25-year period found that the rate of immigration did not affect attitudes while the state of the economy did. Anti-immigration sentiment is affected by individual interests, ideology, and the national economy, but only ideology affects pro-immigration sentiment (Wilkes et al., 2008). This is in line with the study of immigration attitudes through existing sociological theories of *group conflict* (the clashing of groups being rooted in competing economic interests) and *social identity* (the identification with a group as a key element of an individual's sense of who they are) as proposed by Sniderman, Hagendoorn, and Prior (2004). However, this same study found that it was predominantly concerns of culture and national identity rather than economic considerations that resulted in hostile feelings towards ethnic minorities in the Netherlands.

An important distinction is that European countries are not generally classified as traditional 'immigration nations' (Simon & Lynch, 1999), having mostly experienced mass immigration when they were already developed nations. By design, the immigration policy of a country benefits some groups of migrants over others, and in Europe this has historically been driven by labour market needs, later followed by family reunification incentives and political asylum frameworks (Zimmermann et al., 2000). The same study also states that immigrants from countries that are similar to the host country in terms of economic development, the education system, language and culture, assimilate well into the labour market. If humanitarian criteria are used in determining entry into the country, i.e. by concentrating on refugees, successful integration is less likely to take place (Zimmermann et al., 2000). This is reflected by findings from a study on attitudes in Britain (Dustmann & Preston, 2007), stating that the largest single factor of concern is related to the burden of immigrants on the welfare system rather than job displacement, although cultural and racially inclined considerations also influence opinions, except for immigrants from Australia and New-Zealand. This means that policy makers are tasked with finding an equilibrium between economic motives, which would sometimes dictate an increase in the number of migrant workers (Depalo, Faini, & Venturini, 2006), and on the other hand the political and social implications arising from the integration of distinct groups of migrants into the native community.

Classic survey research on migration attitudes has been complemented by analyses of news and media, generally to examine effects of salience and the framing of migration in the media on public opinion. These studies show that not only real-world characteristics, but also their portrayal influences migration sentiment, again varying by place and time. Van Klingeren, Boomgaarden, Vliegenthart, and De Vreese (2015) used Eurobarometer data and human-coded newspaper articles from Denmark and the Netherlands, and found that the effect of real-world immigration figures is limited, while positive news reporting reduces negative attitudes in the Netherlands, but not in Denmark. This was attributed by the authors to the steady presence of the news topic in the Netherlands, which allowed tone to have a distinct effect, whereas media visibility of migration in Denmark was fluctuating. The topic has become increasingly salient since the refugee crisis of 2015, which unshelved existing fears about migration posing a threat to security and the economy (Greussing & Boomgaarden, 2017) but also continued to reveal cross-country differences in the way sentiment appeared in the media (Chan et al., 2020; Heidenreich, Lind, Eberl, & Boomgaarden, 2019). Rather than relying solely on human coders, these studies used newspaper articles leveraged with a sentiment lexicon (Chan et al., 2020) or topic modelling techniques. Interestingly, Greussing and Boomgaarden (2017) found that the framing patterns of quality media versus tabloid media, the latter often providing a more sensationalist and one-sided view of events, became highly similar in times of crisis. This provides a different perspective on a later study (Theorin & Strömbäck, 2020) showing that there are generally limited effects of traditional media consumption on migration attitudes, whereas distinctly anti- or pro-immigration alternative media have a more substantial influence.

2.1 *Leveraging Twitter Data*

More recently, automated sentiment analysis is performed on social media data, including content from popular social network Twitter (Menshikova & Van Tubergen, 2022; Rowe et al., 2021). Twitter is a micro-blogging service that lets users post messages called 'tweets', which can be up to 280 characters long and contain a variety of content such as emojis, images and videos. The mix of anonymous and verified users plus the possibility to interact with specific topics through short identifiers called 'hashtags' results in the platform often serving as a public medium to exchange news, opinions and ideas on a broad set of societal issues, including political events (Freire-Vidal, Graells-Garrido, & Rowe, 2021; Kreis, 2017). Using tweet data together with natural language processing (NLP) techniques enables large scale analysis of mostly unfiltered (Dimitrova, Heidenreich, & Georgiev, 2022) opinions on a topic, investigation of the geographical or temporal distribution of selected expressions, as well as real-time monitoring of changes in sentiment and correlations with real-world events. Research into the framing of migration extends to Twitter, showing that politicians, media and non-governmental organizations dominate the refugee discussion on the platform (Siapera, Boudourides, Lenis, & Suiter, 2018), which may result in the observed prominence of certain frames over others (Dimitrova et al., 2022).

The choice of Twitter as a data source has a number of implications. As opposed to surveys, data is procured directly from a dynamic open space without the need to rely on an accurate phrasing or the correct interpretation of questions to measure public opinion. However, this comes with a different type of bias, namely concerning sampling. Particular groups of the population are over-represented in Twitter data, and significant variability in usage may exist both within and between countries (Rowe et al., 2021). This depends on demographic factors, but also on the activity of users holding particular attitudes on a specific topic: Freire-Vidal et al. (2021) found that users displaying negative immigration sentiment in Chile produced up to 50% more content per user than those with positive attitudes. Nonetheless, Menshikova and Van Tubergen (2022) found that anti-immigrant attitudes on Twitter analysed with the SenticNet and rule-based SentiStrength algorithms mirrored findings from survey literature.

Rowe et al. (2021) studied daily patterns of migration sentiment in Britain over a one-month period and provides key considerations in terms of search strategy, as automated scraping of Twitter data comes at the cost of also capturing noise. Both Rowe et al. (2021) and Inuwa-Dutse, Liptrott, and Korkontzelos (2020) used the VADER model to obtain sentiment scores, each revealing a high prevalence of both negative and positive sentiment which the former attributed to key Brexit events, and the latter to influential Twitter users balancing out high rates of negative sentiment by unverified users. While Inuwa-Dutse et al. (2020) did analyse 800k tweets at a global scale, no tweets in languages other than English appear to have been included. Finally, Chen, Sack, and Alam (2022) propose a publicly available database of multi-lingual tweets on migration collected between 2013 and 2021, and annotated using a fine-tuned BERT model for sentiment analysis. It includes over 27k tweets from Italy and 15k tweets from the Netherlands besides nine other European languages, and the temporal distribution of sentiment labels for all 11 countries shows a consistently high prevalence of neutral tweets, while the proportions of negative versus positive sentiment fluctuate throughout the period of study. A possible limitation of the study is revealed by the search strategy: keywords (Chen et al., 2022) for data collection were gathered based on the 50 most similar terms to the seed words *refugee* and *immigration* according to Google News and fastText, with some languages showing a high presence of terms with negative connotation such as 'jihadist', 'neo-nazi', and 'terrorist' in Polish.

3 EXPERIMENTAL SETUP

This section consist of two components. First, we provide a description of the data that was used for this analysis. Second, we present the chosen algorithms and outline the research methodology. An illustration of the research workflow can be found in Figure 1.

3.1 *Data*

A Python script was developed to collect tweets from leading news and media outlets via the open Twitter Application Programming Interface (API). We relied on the full-archive endpoint available with Academic Research access (`https://developer.twitter.com/en/products/twitter-api/academic-research`) and used the Tweepy library for Python as the wrapper to assemble a corpus of tweets for both countries. The script collected all tweets from selected outlets posted between 2022-01-01 and 2022-11-11, on the condition that there were at minimum 10 replies to the original tweet to limit the number of requests to the API. For each of the collected tweets, the full reply thread was also collected. The outlets in the sample include the Twitter accounts of public news broadcaster NOS and online news portal NU.nl for the Netherlands. The NU.nl portal can be accessed through its website or mobile app and is owned by DPG Media Group, the publisher that also owns nationwide newspapers *Algemeen Dagblad*, *De Volkskrant*, and *Trouw* as well as multiple regional papers with high local readership numbers. Articles on the portal are based on the latest news articles from these publications, but NU.nl also produces original content related to current affairs. For Italy, the accounts of newspapers *La Repubblica*, *Corriere della Sera*, and *Il Fatto Quotidiano* were included, as well as those of several larger and smaller news formats owned by the RAI state media network. These include the Twitter accounts of its main news program *Rai News* (@RaiNews) and press office RAI Ufficio Stampa (@Raiofficialnews), the news bulletins from its public TV channels Rai 1, Rai 2 and Rai 3 (@Tg1Rai, @tg2rai, @Tg3web) and finally, of several political or current affairs talk shows: *Agora* (@agorarai), *Porta a Porta* (@RaiPortaaPorta), and *Report* (@reportrai3). They have all been aggregated under the label 'RAI Group' for this analysis.

Table 1: Search terms used for tweet data selection.

Country	**Search terms**
The Netherlands	migratie, migrant, vluchteling, statushouder, veiligelander, asiel, 'Ter Apel'
Italy	migrazione, migrant*, immigrat*, profug*, rifugiat*, clandestini, asilo, sbarco, sbarchi, 'Geo Barents', Humanity, 'Ocean Viking'

A set of language-specific search terms was then defined to filter the data and retain only the content relevant to the context of migration. The search terms are approximately equivalent in both languages to ensure a similar corpus makeup. The exception are two topics that define the current reality of each country: the Netherlands dealing with shelter and housing capacity problems for asylum seekers, most notably in the central asylum registration centre located in Ter Apel, and Italy being the entry point for the largest share of all irregular migrants to Europe, who arrive predominantly via sea through the so-called Central Mediterranean migratory route (Frontex, n.d.). All search terms are listed in Table 1, and an overview of the tweet data samples after filtering is provided in Table 2.

To limit pollution of our samples with irrelevant content, additional search strings were formulated to filter out unrelated news tweets associated with specific search terms. For example, 'asilo' (*asylum*) also yields results related to multiple tragic events at a children's daycare facility, commonly called 'asilo nido' or simply 'asilo' in Italian. Other frequent noise topics include the controversy surrounding foreign construction workers in Qatar for the FIFA World Cup in men's football, and internally displaced civilians due to Russian air strikes on cities in Ukraine. Due to the limited number of news tweets, topic identification was accomplished manually by the author, i.e. suitable search strings to remove noise were formulated upon inspection of the filtered data, rather than opting for an automated topic modelling approach. As the body of comments was significantly larger and linked to the news tweets, no noise topic removal was performed on the reply samples. To achieve a 'clean' text format, we removed hyperlinks, the '#' symbol from the beginning of hashtags, names of journalists in brackets, and mentions of the type '@user' from all tweets using regular expression patterns. Lastly, a test set ($n = 200$) was randomly sampled from each of the news and reply subsets per country to be hand-labelled by the author, for the purpose of evaluating model performance on our specific task.

Table 2: Data volumes after filtering and removing noise topics.

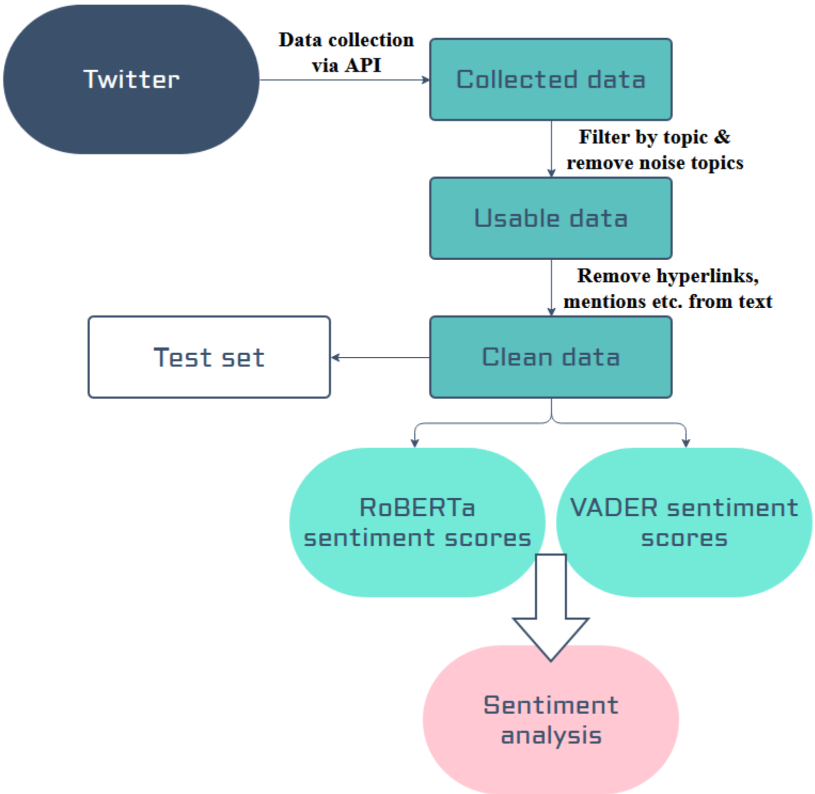
		Tweet Counts	
Country	Outlets	News	Replies
The Netherlands	NOS (@NOS)	309	27,255
	NU.nl (@NUnl)	303	27,115
	Total NL	**612**	**54,370**
Italy	La Repubblica (@repubblica)	76	11,520
	RAI Group	50	800
	Il Fatto Quotidiano (@fattoquotidiano)	35	1,384
	Corriere della Sera (@Corriere)	34	1,249
	Total IT	**195**	**14,953**

3.2 *Method / Models*

The first model we applied is the *Valence Aware Dictionary and sEntiment Reasoner* or VADER, a rule-based sentiment scoring tool that is specifically attuned to textual expressions on social media. This is due in part to its ability to grasp features like internet slang (such as 'meh'), many acronyms (such as 'LOL'), capitalisation, punctuation, classic emoticons, and even emojis as having a distinct boosting or nuancing effect on the sentiment polarity of a text message (Hutto & Gilbert, 2014). Furthermore, it accurately handles negations and mixed-intensity phrases (e.g. *'As generously as we treat Ukrainian refugees, our treatment of Syrian and African refugees is bad while they too are fleeing war and violence'*). On sets of labelled tweets, reviews, and both news articles and editorials processed at the sentence-level, VADER outperformed both trained human raters and other sentiment analysis tools (Hutto & Gilbert, 2014). Valence scores (expressing both sentiment polarity and intensity) are calculated based on the words occurring in VADER's lexicon of word tokens: for example, the word 'okay' carries a valence of +0.9 while 'horrible' and 'great' have -2.5 and +3.1 respectively. Scores are summed across the entire text snippet, then normalised to fall on a scale of -1 to +1. The output is a weighted, normalised compound score within this range, as well as a measure of the proportions of the input text that fall into the different classes of sentiment polarity. The aforementioned mixed-intensity phrase, for example, is scored as follows: {'neg': 0.286, 'neu': 0.49, 'pos': 0.224, 'compound': -0.5267}. Although it has been recognised as an accurate scoring tool in sentiment analysis studies on similar data (Inuwa-Dutse et al., 2020; Rowe et al., 2021), it is worth noting that there are certain drawbacks to the rule-based approach. For instance, accurately detecting sarcasm remains difficult (Rowe et al., 2021) as it tends to be of a non-literal nature, and typically requires insight into the circumstances of the expression. A multilingual VADER sentiment package for Python is available in the PyPi repository (`https://pypi.org/project/vader-multi/`), integrating the Google Translate API to allow for sentiment analysis on non-English texts. This model was selected as a baseline for our approach.

The second model employed for our task is a BERT-based deep learning language model. BERT is an acronym of *Bidirectional Encoder Representations from Transformers*, an encoder-decoder NLP model that is itself based on the Transformer architecture first proposed by Vaswani et al. (2017). It follows the conventional neural network structure for sequential input data such as language (i.e. sequences of words), consisting of multiple encoder and decoder layers. A key element of the Transformer is the multi-headed self-attention mechanism in the encoder layers, which allows it to retain important contextual information from the input sequence. This is done by attaching weights to individual word tokens with respect to the context that they occur in. Given sufficient training data, the network is able to gather word association knowledge (which words occur frequently together with a given word) and pick up on features such as positional information (where a given word occurs in a sentence) by learning the optimal values for the self-attention weights to produce an accurate output. The BERT model developed by Devlin, Chang, Lee, and Toutanova (2018)

Figure 1: An illustration of the research workflow from data collection to analysis.



implemented this mechanism to train on two language processing tasks in a self-supervised manner, using batches of text data from a large book corpus and English Wikipedia articles. RoBERTa (Liu et al., 2019) is an optimised variant of the BERT architecture, achieving better performance by keeping just one pre-training task (*masked language modelling*, i.e. predicting the most probable word when masked in a sentence) and slight tuning of key hyperparameters, such as batch size. Thus, the conventional workflow for a BERT-based model consists of two components: pre-training on a large corpus of text data, and fine-tuning to improve accuracy on one's intended NLP task. In fine-tuning, one or more fully connected layers are typically added on top of the final encoder layer of a trained model (Rogers, Kovaleva, & Rumshisky, 2020) to map the features within its previous layers to the output classes.

The choice of a BERT-based model as the gold standard was motivated by its performance exceeding that of every other model currently available (Liu et al., 2019), and its predisposition for transfer learning. Pre-training a language model is often computationally expensive, and requires large amounts of language data not at our disposal. We therefore made use of the Transformers library for Python by Hugging Face (Wolf et al., 2019) that provides an open-source pipeline to implement pre-trained Transformer architectures for a variety of NLP tasks. We specifically selected a cross-lingual model based on RoBERTa (Barbieri, Anke, & Camacho-Collados, 2021), that was pre-trained on 198M tweets in more than 30 languages and fine-tuned for sentiment analysis on 8 of these languages, including Italian. The model is used with its own tokeniser that maps input tokens from our tweets against its acquired vocabulary, and outputs an array with a softmax-transformed probability score between 0 and 1 for each polarity class, the larger being the final prediction label ('negative', 'neutral' or 'positive'). Using the pre-trained XLM-RoBERTa model by Barbieri et al. (2021), the same mixed-intensity phrase mentioned above yields the following output: {'label': 'negative', 'score': 0.8954832553863525}.

A comparison between the polarity labels assigned by VADER and XLM-RoBERTa to the tweets in our evaluation set will provide an answer to sub-question *Q.1* regarding each model's performance on our task, in addition to illustrating the robustness of our approach. Guidelines for annotation and the distribution of classes in the evaluation set can be found in Appendix A. It should be noted that the numeric values of the scores are not directly comparable between models. The weighted compound score by VADER inherently expresses polarity through a negative or positive sign, while RoBERTa simply outputs the probability scores for each class in the final layer of the neural network, together with the assigned label. By visualising the distribution of model outputs for the tweets in each of our outlet, user and country samples, we will be able to answer sub-questions *Q.2* and *Q.3* about the degree of polarisation in selected user groups.

4 RESULTS

In this section, the results of the model performance comparison will be presented, followed by the findings of the sentiment analysis as proposed in Section 1.4.

4.1 *Model Evaluation*

Table 3 shows the results of the model evaluation. The weighted F_1 score is reported, as there is a significant class imbalance with regard to sentiment polarity labels in the evaluation set (see Appendix A). A first key observation is that both models appear to perform significantly better on the news tweets than on the replies. As the dataset used for this study contains approximately 8 times more reply tweets than news tweets, the 'Total' score is less representative for overall performance than the F_1 scores for 'Total News' and 'Total Replies' separately. Here, too, it should be noted that the original dataset contains over 3.5 times more Dutch than Italian tweets, meaning that the performance on our full sets is likely to differ from the scores reported here.

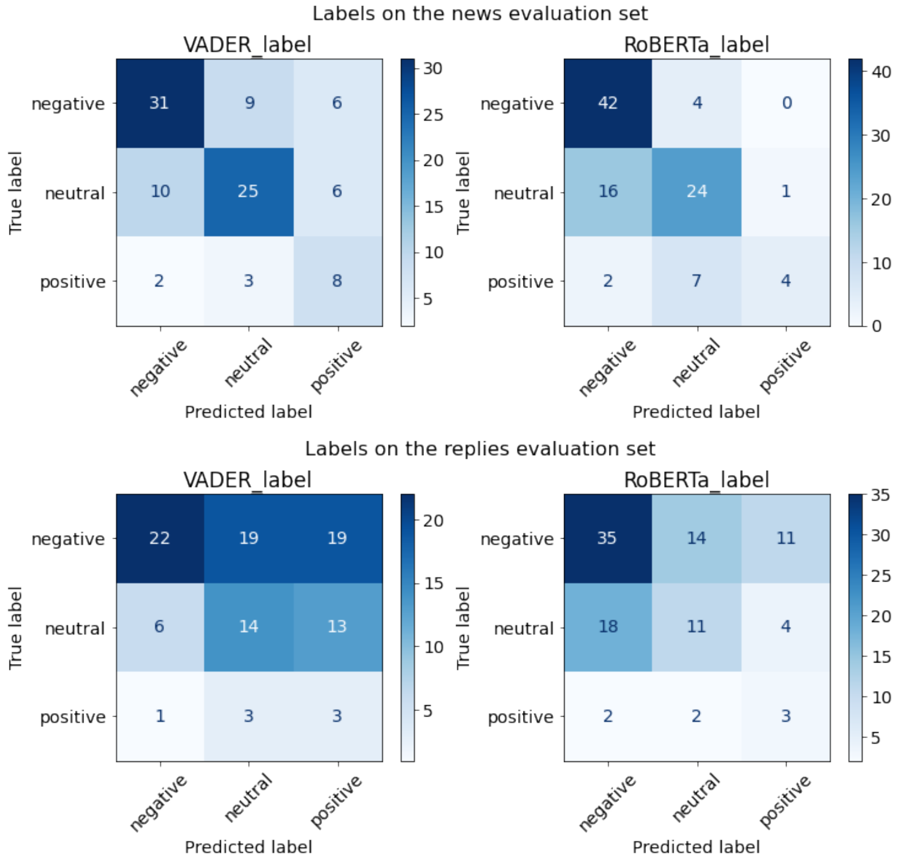
Table 3: Model performance on the evaluation set. F_1 scores are weighted scores.

		F_1 score	
Subset	Language	VADER	RoBERTa
News	Dutch ($n = 50$)	0.589	0.748
	Italian ($n = 50$)	0.702	0.580
	Total News	**0.646**	**0.681**
Replies	Dutch ($n = 50$)	0.378	0.477
	Italian ($n = 50$)	0.503	0.531
	Total Replies	**0.441**	**0.503**
	Total ($n = 200$)	**0.541**	**0.592**

In addition to the user category, performance appears to vary depending on the language of the tweet. F_1 scores are generally higher for the Italian subsets, with the exception of RoBERTa obtaining its highest F_1 score of 0.748 on the Dutch news tweets. VADER shows the opposite, with an F_1 score of 0.702 on the Italian news tweets representing its highest performance among the different subsets.

The confusion matrix in Figure 2 shows that both models do best on the news data, with neutral and positive news tweets proving slightly more problematic for RoBERTa than for VADER. For the replies, VADER incorrectly labels negatively valenced tweets as neutral or even positive comments. RoBERTa's classification of the replies shows a similar but more accurate pattern, although the model appears to be biased towards assigning a negative label, while VADER shows a tendency for evenly balanced ratings. The difference between the models is especially apparent in the neutral and positive classes. It should however be noted that the hand-annotated sets were rather small ($n = 100$ per user category), which is perhaps too small to be representative of the full dataset. Figure 3 shows RoBERTa performing better than VADER on Dutch tweets while the difference for Italian is less obvious, but again mostly visible in the neutral and positive class. Similarly to the results by user category, VADER shows

Figure 2: Predicted class labels by both models on the news and reply sets. True labels represent human annotation.

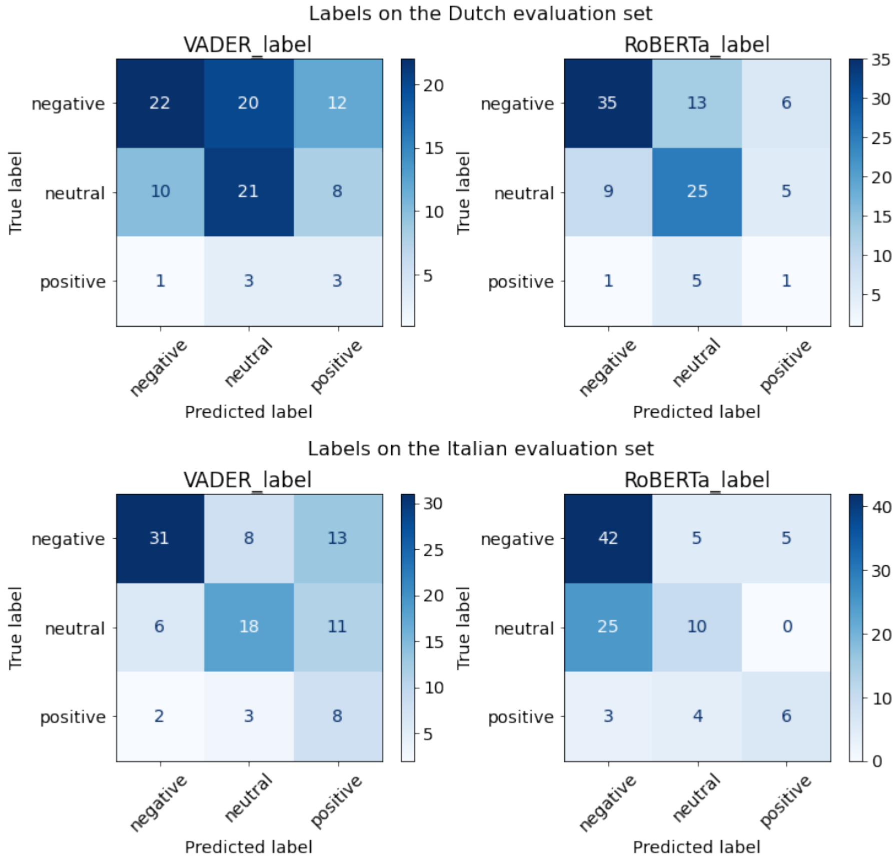


a more evenly distributed set of assigned class labels while RoBERTa leans towards a negative classification.

4.2 *Sentiment Analysis Results by Country*

Figure 4 presents the scoring results on the news dataset. As described in Section 3.2, output scores differ per model with RoBERTa requiring the label to be included for a comparative analysis. Both models detect a higher frequency of neutral tweets in the Dutch news sample, while the Italian news sample contains relatively more positive and negative tweets with respect to the total amount of collected tweets. However, RoBERTa paints a different picture: not only does the negative class severely outweigh the positive class in the Italian data, it is also significantly larger than the volume of neutral news tweets. In addition, the skew displayed by the curve of the negative class indicates that a large portion of these tweets

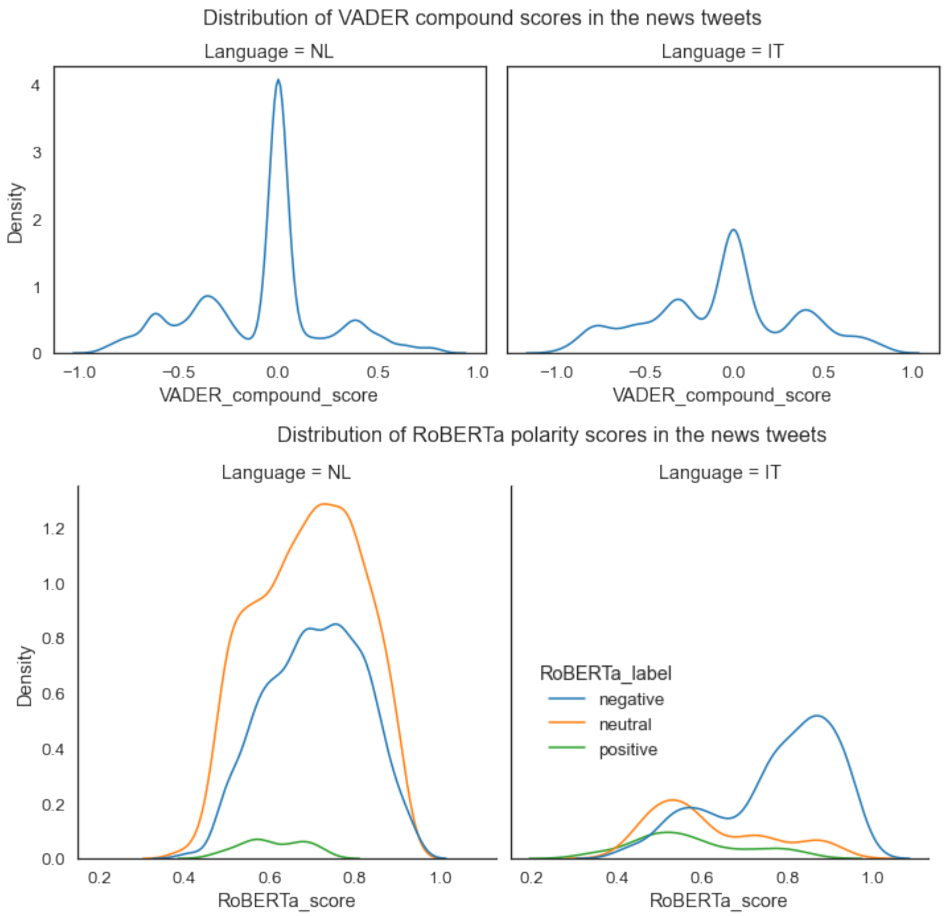
Figure 3: Predicted class labels by both models on the Dutch and Italian sets. True labels represent human annotation.



lie in the higher range of scores, meaning that they are in fact intensely negative. Both Italy and the Netherlands show relatively little positively valenced news, although the proportion of positive tweets in the Italian set is clearly higher than in the Dutch news data.

Figure 5 shows the results on the sets of replies. Contrary to the distribution of news tweets, the reply curves are very similar between Italy and the Netherlands, with negative sentiment being the most prominent in both countries according to RoBERTa. Both models agree that the proportion of neutral sentiment is higher in the Dutch set, but interesting here is also the higher density of positive replies compared to the Italian set, which is especially obvious in the distribution of scores by RoBERTa. This is different from the picture presented by the news tweets, where positive tweets made up a larger share of the Italian news set than the Dutch set. Finally, there is also a noticeable peak in the distribution of neutral tweets by RoBERTa, caused by tweets with a low polarity score of

Figure 4: VADER weighted compound scores and RoBERTa scores on the news tweets.



approximately 0.4. This indicates the presence of some positive or negative sentiment in the neutral replies, which could perhaps be understood as the presence of 'both' or mixed-intensity sentiments among regular users.

4.3 *Sentiment Analysis Results by Outlet*

Figure 6 shows the distribution of sentiment in tweets by Dutch outlets NU.nl and NOS, while Figure 7 shows the replies. The density of negative tweets by NU.nl approaches that of their neutral tweets, while in the NOS sample neutral news is dominant. Positive tweets are a minority class in both samples. The distribution of replies is highly similar between outlets, with replies to NOS showing perhaps a slightly higher incidence of positively valenced tweets.

Figure 5: VADER weighted compound scores and RoBERTa scores on the replies.

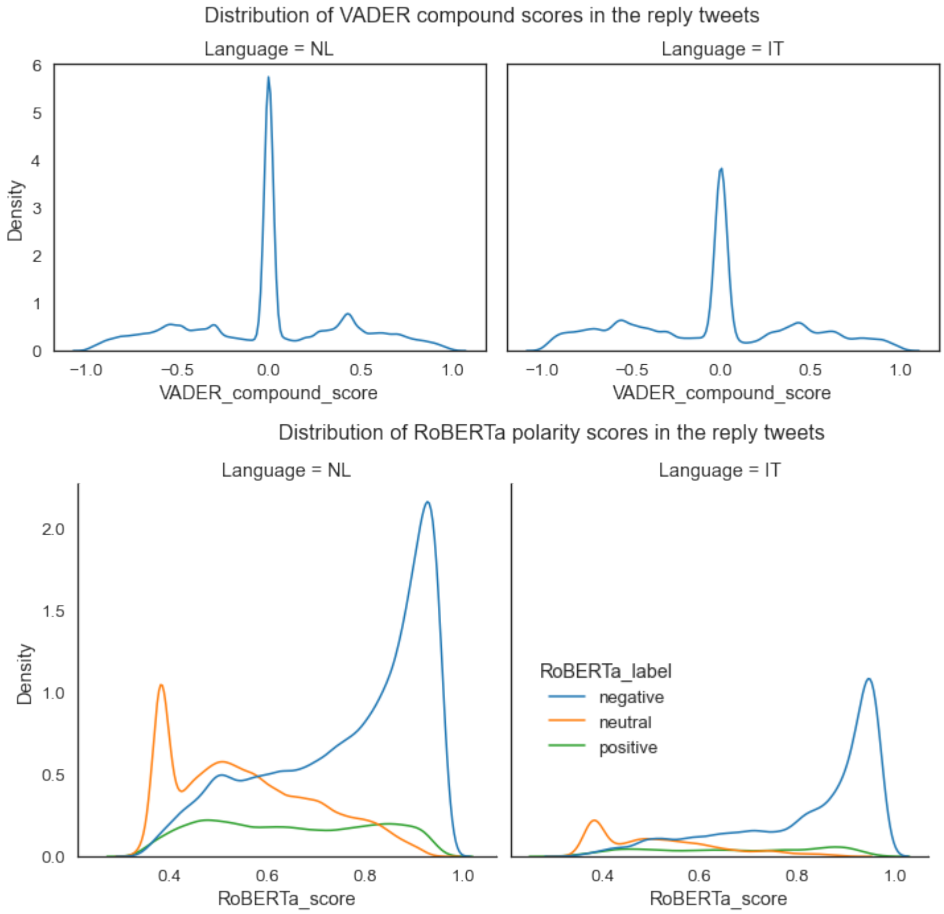


Figure 8 shows the distribution of sentiment in tweets by Italian outlets RAI, *La Repubblica*, *Corriere* and *Il Fatto Quotidiano*. Negative sentiment is the most prominent in all outlets, with *Il Fatto Quotidiano* clearly showing the highest density of positive news tweets among the four. Figure 9 shows that negative sentiment also dominates the reply samples for each Italian outlet.

Results by VADER show minor variation between outlets, with findings mostly corresponding to the observations from the distributions by RoBERTa presented above. For reasons of coherence, these Figures have been included in Appendix B.

Figure 6: Distribution of RoBERTa scores by outlet in news_nl.

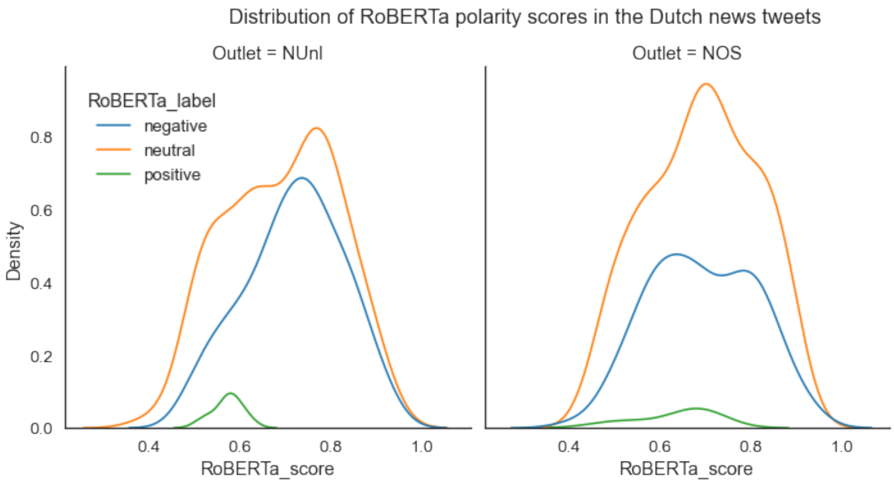
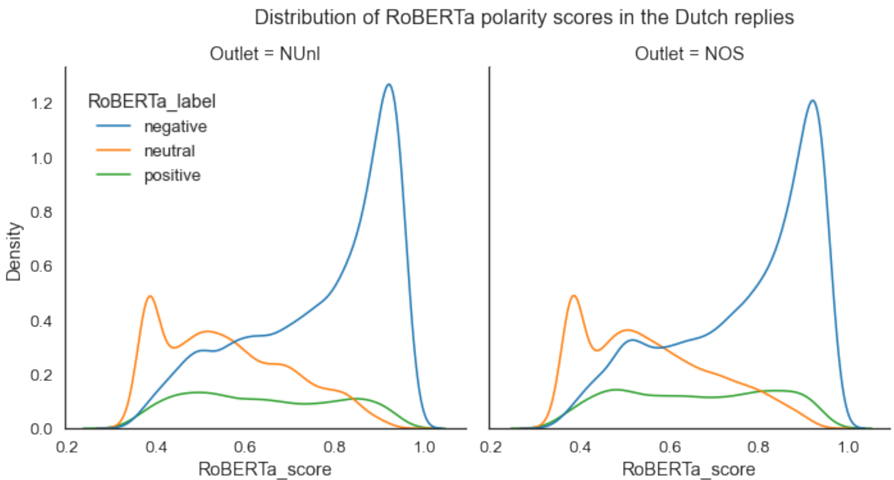


Figure 7: Distribution of RoBERTa scores by outlet in replies_nl.

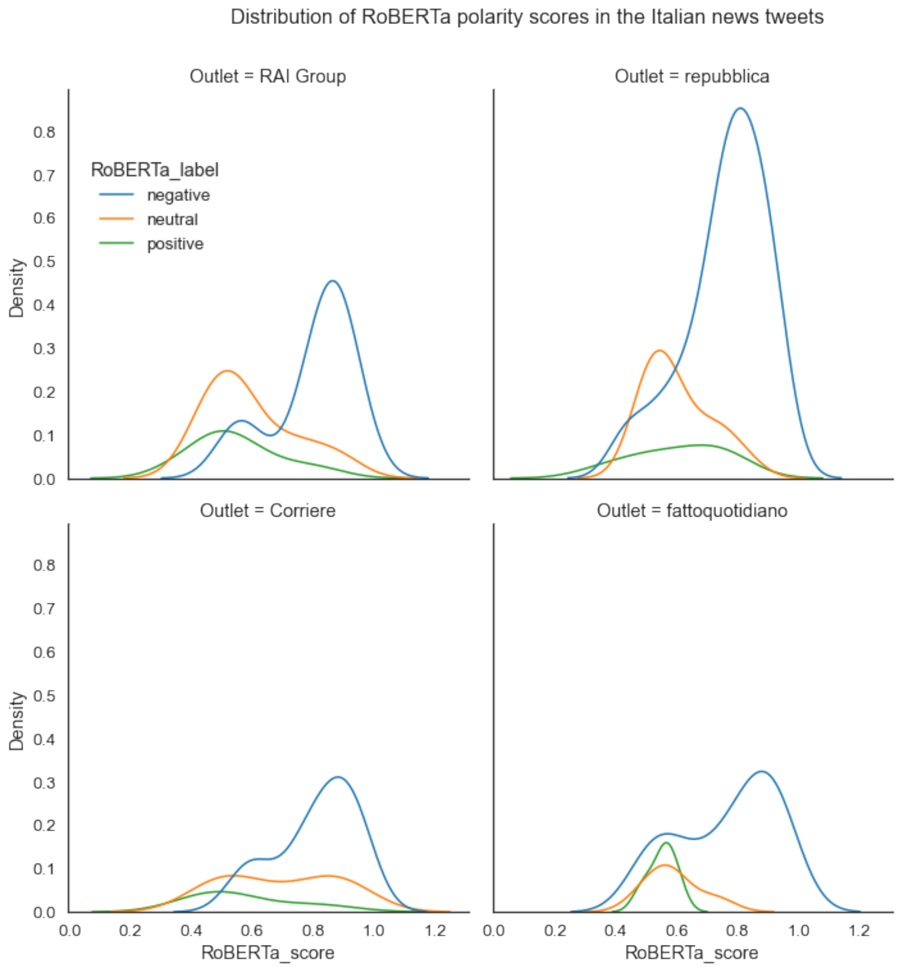


4.4 *Post-Hoc Analysis*

As both Theorin and Strömbäck (2020) and Meltzer et al. (2021) proposed that increased coverage of immigration raised its perceived importance as a problem, with Meltzer et al. (2021) also stating that this in turn increased support for anti-immigrant parties, it is perhaps more fruitful to explore variation in the distribution of news tweets over time.

Comparing Figure 10 and Figure 11 suggests that it is not the tone of the news message that determines sentiment polarity in the replies. A peak in neutral news tweets in the Netherlands around September 2022 that was over 1.5 times the size of the negative peak in the same period, resulted in

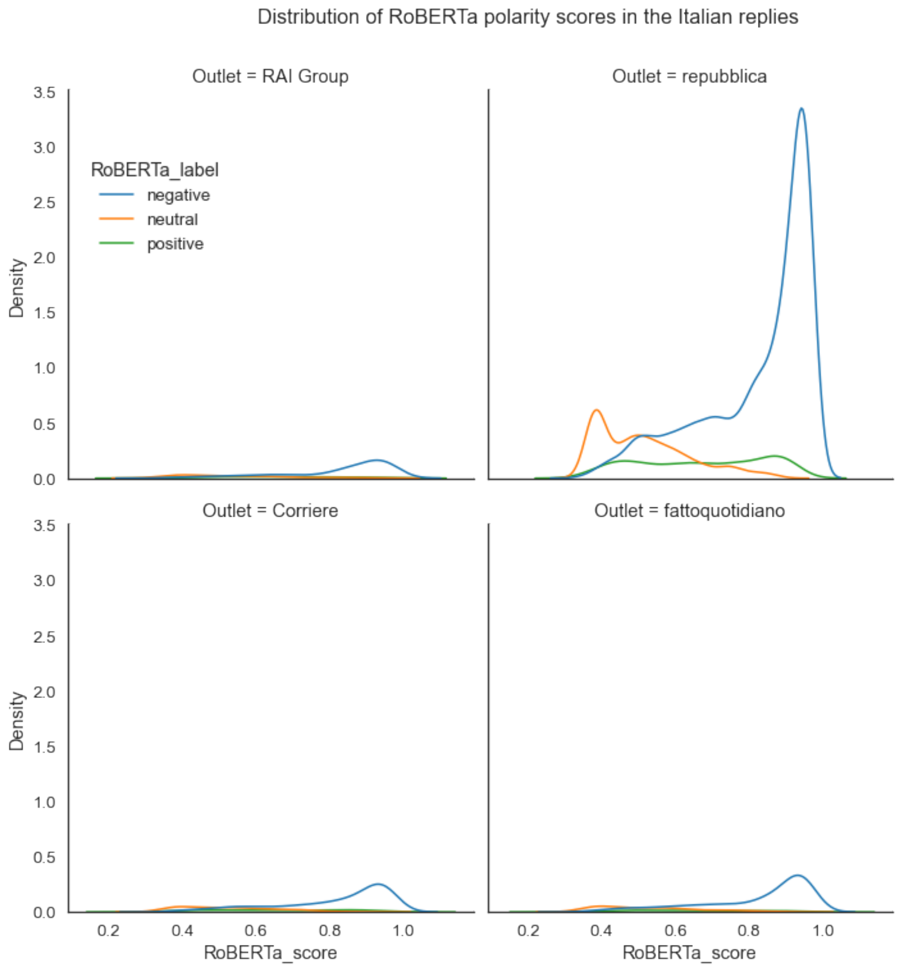
Figure 8: Distribution of RoBERTa scores by outlet in news_it.



a reaction of opposite sentiment levels, with negative sentiment more than twice as frequent as neutral responses. Moreover, while almost no positive news tweets were posted during this time, there is also a peak in positive sentiment in the replies. Another key observation is that the Italian tweets were mostly concentrated around November 2022. This reveals an uneven composition of the country samples, and likely means that the observed dominance of negative sentiment in the Italian samples is not generalisable to other time periods.

In addition to the above, Figure 11 reveals a number of timestamps with peak activity in the reply samples. To further investigate these data points, the five news tweets that evoked the highest response in each country have been included in Appendix C.

Figure 9: Distribution of RoBERTa scores by outlet in replies_it.

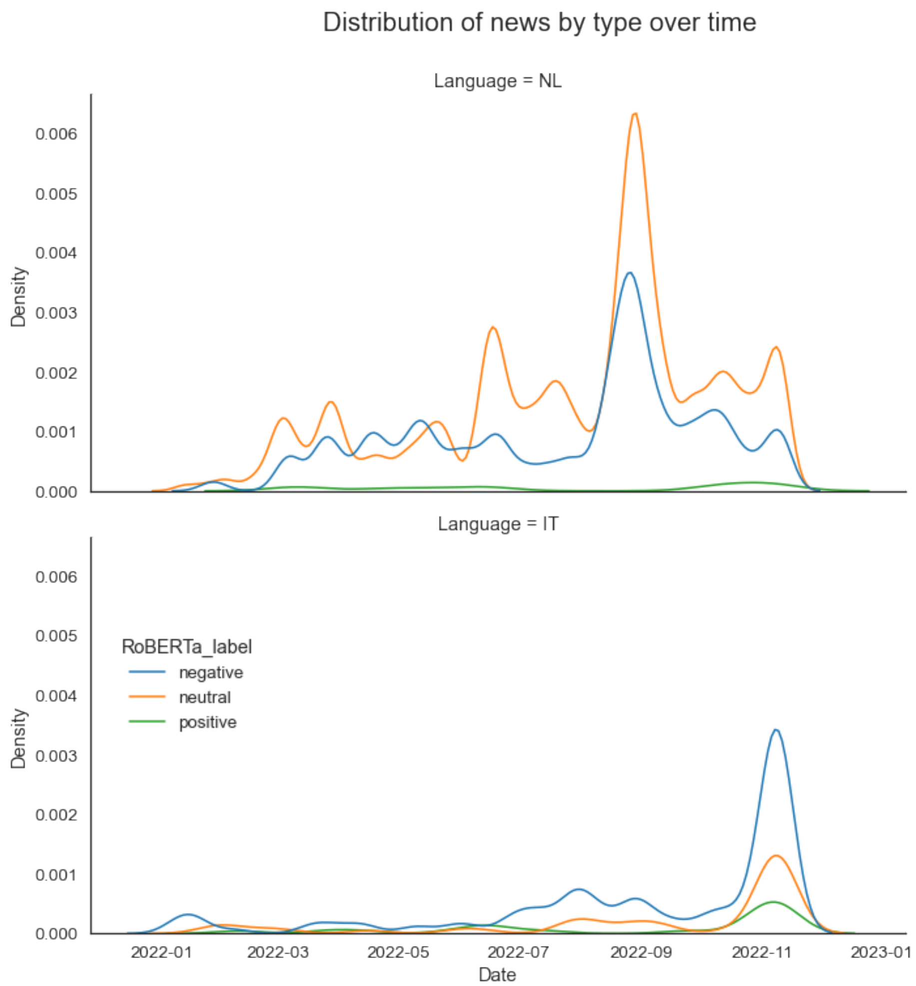


5 DISCUSSION

5.1 *Error Analysis of Model Performance*

The results of the model comparison proposed in sub-question *Q.1* confirmed the hypothesis regarding the performance of a pre-trained BERT-based model surpassing that of the rule-based VADER sentiment analyser. However, a number of critical remarks can be made with regard to the chosen evaluation method. Firstly, creating a human-annotated 'gold standard' for supervised learning generally involves employing at least two independent human raters (Dimitrova et al., 2022; Menshikova & Van Tubergen, 2022). While no recommendations exist for the minimum required size of the labelled sample, at least to the knowledge of the author, it could also be argued that a sample size of $n = 50$ per category is not sufficient

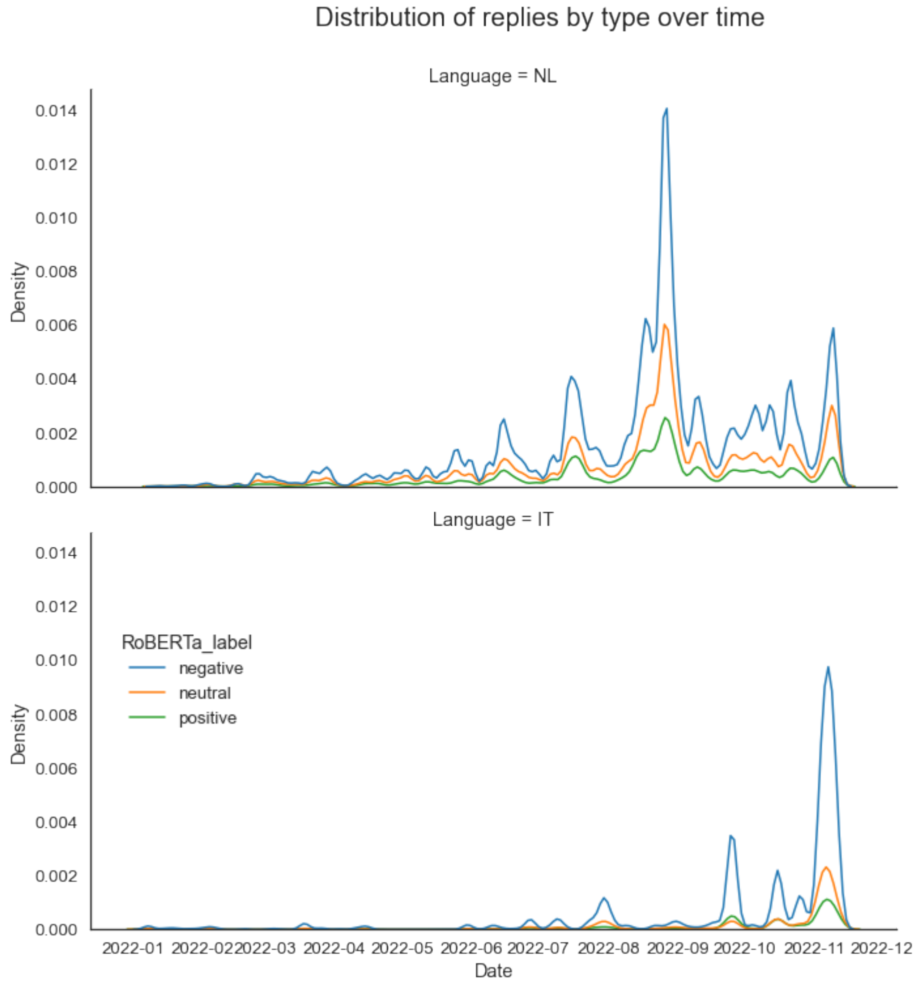
Figure 10: The frequency of news tweets in the dataset over time.



to make reliable inferences about the accuracy of the model on different sample types. Secondly, although great care was taken to ensure accurate annotation (see Appendix A) of the test set, the low model performance compared to the literature (Hutto & Gilbert, 2014; Liu et al., 2019) could at least in part be due to human error, especially considering the non-native proficiency of the author in Italian. Interestingly, sentiment fine-tuning for XLM-RoBERTa (Barbieri et al., 2021) was done on Italian but not on Dutch, which means that although the model can be used for Dutch, it was anticipated to do better on Italian data.

Another plausible explanation for the higher-than-expected error rate is the fact that Twitter content varies between user groups. Tweets from reputable news sources are presumably less likely to contain spelling errors, ridicule, or sarcasm, which are factors that are known to complicate

Figure 11: The frequency of news tweets in the dataset over time.



the task of an NLP model, as noted in part by Rowe et al. (2021). This matter can partially be resolved with better data-preprocessing techniques. While no lowercasing, stemming or lemmatisation was performed due to both models working best on an input at the sentence level, removing excess punctuation or applying a spell checker such as the `.correct()` function from the TextBlob library for Python might have resulted in better performance on the replies. Unfortunately, this option is currently only available for input text in English, French and German. The tendency for VADER to underestimate the number of negative tweets compared to RoBERTa or a human annotator, especially in the reply subset, could be due to the model misinterpreting ridicule (such as 'HAHAHA' or crying-with-laughter emojis) as expressions of joy. This also applies to RoBERTa, as the model may be more context-aware, but the news tweet is not accessible for either model during the scoring of replies. The author labelling the reply tweets while being aware of the context, i.e. the news tweet that provoked the response, may be comparable to the experience of a reader on Twitter but constitutes an 'unfair' disadvantage over the model, which is a flaw in the design of the evaluation method.

5.2 *Interpretation of Sentiment Analysis Results*

Findings from Figure 4 confirm the hypothesis proposed to sub-question *Q.2*, stating that migration-related tweets in Italy reflect a higher degree of polarisation than tweets in the Netherlands. This appears to be the case for the findings from the news samples. However, the same does not hold true for the replies, as shown in Figure 5. While neutral tweets dominate the Dutch news set, both positive and negative comments are more frequent in the Dutch than in the Italian set of replies. This would mean that there is no relation between polarisation of the public debate and the political climate of a country, but instead that higher proportions of negative sentiment in the population correlate with political preferences. In addition, the Figures suggest that the tone of the news does not have a systematic effect on the emotional state of the audience, or that there are other factors at play, suppressing any effects of tone. To explore this effect, one could proceed to sample different subsets of news tweets by score and analyse the sentiment in the corresponding replies.

A flaw in the data cleaning process may have contributed to the prominence of neutral tweets in all reply samples. As links or mentions cannot be interpreted by either model, they were replaced with an empty string, which resulted in replies that contained only these elements inflating data volumes without contributing to the analysis. It is therefore advisable to replace these with a null value and then remove them from the sample, or perhaps remove replies with a length under two characters entirely, so that only meaningful content remains.

With regard to the difference between outlets as proposed in sub-question *Q.3*, no clear pattern can be observed, besides the replies being predominantly negative for all outlets. Findings from Section 4.2 and Section 4.3 mostly refute our hypothesis that sentiment in media tweets should remain relatively neutral. Italian media outlets in particular show a tendency to tweet negatively about migration-related topics, although results by VADER disprove these findings (see Appendix B) and this is likely to have been caused by the uneven distribution over time of our Italian set of tweets. As far as our sample is concerned, neutral reporting of migration does dominate the Dutch media landscape but here, too, the proportion of negative tweets is relatively high for both outlets. The hypothesis that tweets by the public carry more strongly positive and negative sentiment than news tweets only holds true for the Dutch samples, but a lack of reply data for three out of four Italian outlets means that no robust inferences can be made with regard to this comparison. Figure 9 and Figure 15 in Appendix B reveal that a number of the larger Italian news outlets do not in fact generate high rates of interaction per post, as the contributions to the body of replies by readers of *Corriere della Sera*, *Il Fatto Quotidiano* and the accounts under RAI Group on Twitter are clearly very small.

Finally, the assumption that readers of more biased outlets express stronger sentiment in their own tweets cannot be reliably confirmed or refuted based on our findings. The Dutch data suggests no relation, as shown in Figure 6 and Figure 7, while the Italian data is too limited to draw any conclusions in this regard (see Figure 8 and Figure 9). Post-hoc analysis suggests that not the tone, but rather the salience of the topic in the media predicts the emergence of negative migration sentiment in Twitter users, in addition to specific real-world events. In line with classic sociological theories and previous studies, tweets that provoked the most response were related to *realistic group conflict* (e.g. competition for resources, welfare concerns) as well as valenced real-world events and statements by influential figures, supporting findings from similar studies by Freire-Vidal et al. (2021) and Menshikova and Van Tubergen (2022).

5.3 *Limitations and Further Work*

The results of this analysis show that computational sentiment analysis models can assess the emotional tone of Twitter data to varying degrees. In addition to the design improvements suggested above, better results could likely be obtained with RoBERTa after fine-tuning on large amounts of annotated data, such as the tweets in MGKB (Chen et al., 2022). With regard to VADER, it may prove worthwhile to alter or extend the built-in lexicon in the source code, and thus tailor it to the intended use case.

The question whether migration-related sentiment varies systematically between outlets and their audience remains unanswered. Regardless of the lack of sufficient data in the Italian reply set, this effect is perhaps more reliably studied through robust multilevel statistical models, such as by Van Klingeren et al. (2015) or Menshikova and Van Tubergen (2022). Nonetheless, Twitter data provides a rich source of information to complement traditional survey research, allowing for exploration beyond the textual expressions, e.g. through network analysis (Menshikova & Van Tubergen, 2022) and the impact of specific real-world events (Siapera et al., 2018). Some authors (Van Klingeren et al., 2015) argue that the causal mechanism between media and public attitudes is reversed, and it is in fact the media that pick up on public tendencies. This scenario can undoubtedly be investigated through setups similar to the one in this thesis.

A final limitation concerning the majority of studies cited for this work is that the focus lies on addressing misrepresentations and other frame-related drivers of negative sentiment, but rarely on real-world factors that cause anti-immigration attitudes. While public perceptions are intensely monitored (European Commission, n.d.) at a European level, these insights rarely translate to decisions that could lead to a more sustainable and thus better-received policy at the national level. Regarding public opinion, much attention is devoted to negative sentiment and anxiety about migration, while 'toxic positivity' or the dismissal of legitimate concerns as racism, xenophobia, or intolerance has equally detrimental results, fueling extremism (Papademetriou & Banulescu-Bogdan, 2016). Meanwhile, a relatively unexplored strategy to reduce conflict and opposition is a heightened emphasis on the assimilation of new communities, as proposed in Kaufmann (2019).

6 CONCLUSION

This thesis presents a machine-learning approach to analyse migration-related sentiment on Twitter in two different countries. It aims to answer the following research question: *'To what extent can a rule-based model versus a neural network language model identify polarisation in tweets about migration?'*. The results suggest that computational sentiment analysis models can accurately assess the emotional tone of Twitter data, with the BERT-based neural network model showing the most promising results. However, performance is to a large extent dependent on the quality of the input data.

A relationship between media tone and a polarised public discussion could not be proven, although the mere salience of the topic in the media is shown to elicit mixed responses, with negative sentiment being the most prominent. This is in accordance with findings from related studies. Implications for society and policymakers are the persistent relevance of the topic in public debate, as well as the demonstrated effects on political developments if no sustainable solutions are achieved. Following the recommendations by Papademetriou and Banulescu-Bogdan (2016), the focus should lie on targeting irregular migration and striking a balance between accommodation, adaption, and restriction to ensure successful integration, while also meeting humanitarian obligations. Overall, the findings provide evidence that hypotheses from existing sociological theories also effectively explain computer-identified migration sentiment in social media expressions.

REFERENCES

Barbieri, F., Anke, L. E., & Camacho-Collados, J. (2021). XLM-T: A multilingual language model toolkit for Twitter. *arXiv preprint arXiv:2104.12250*.

Chan, C.-H., Wessler, H., Rinke, E. M., Welbers, K., Atteveldt, W. v., & Althaus, S. (2020). How combining terrorism, Muslim, and refugee topics drives emotional tone in online news: A six-country cross-cultural sentiment analysis. *International Journal of Communication: IJoC*, *14*, 3569–3594.

Chen, Y., Sack, H., & Alam, M. (2022). Analyzing social media for measuring public attitudes toward controversies and their driving factors: a case study of migration. *Social network analysis and mining*, *12*(1), 1–27.

Dennison, J., & Geddes, A. (2019). A rising tide? The salience of immigration and the rise of anti-immigration political parties in Western Europe. *The political quarterly*, *90*(1), 107–116.

Dennison, J., & Geddes, A. (2021). Thinking globally about attitudes to immigration: Concerns about social conflict, economic competition and cultural threat. *The Political Quarterly*, *92*(3), 541–551.

Depalo, D., Faini, R., & Venturini, A. (2006). The social assimilation of immigrants. *CEPR Discussion Paper Series*.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dimitrova, D., Heidenreich, T., & Georgiev, T. A. (2022). The relationship between humanitarian NGO communication and user engagement on Twitter. *New Media & Society*.

Dustmann, C., & Preston, I. P. (2007). Racial and economic factors in attitudes to immigration. *The BE Journal of Economic Analysis & Policy*, *7*(1).

European Asylum Support Office. (2015). *Latest asylum trends – 2015 overview.* (https://euaa.europa.eu/sites/default/files/public/LatestAsylumTrends20151.pdf. Retrieved on: 2022-12-07.)

European Commission. (n.d.). *Public perceptions of migration and integration.* (https://knowledge4policy.ec.europa.eu/foresight/topic/increasing-significance-migration/integration-strategies-european-commission-proposal-migration-border-management_en. Retrieved on: 2022-12-27.)

Freire-Vidal, Y., Graells-Garrido, E., & Rowe, F. (2021). A framework to understand attitudes towards immigration through Twitter. *Applied Sciences*, *11*(20), 9689.

Frontex. (n.d.). *Migratory routes.* (https://frontex.europa.eu/we-know/migratory-routes/central-mediterranean-route/. Retrieved on: 2022-12-19.)

Greussing, E., & Boomgaarden, H. (2017). Shifting the refugee narrative? An automated frame analysis of Europe's 2015 refugee crisis. *Journal of Ethnic and Migration Studies*, *43*(11), 1749–1774.

Heidenreich, T., Eberl, J.-M., Lind, F., & Boomgaarden, H. (2020). Political migration discourses on social media: A comparative perspective on visibility and sentiment across political Facebook accounts in Europe. *Journal of Ethnic and Migration Studies*, *46*(7), 1261–1280.

Heidenreich, T., Lind, F., Eberl, J.-M., & Boomgaarden, H. G. (2019). Media framing dynamics of the 'European refugee crisis': A comparative topic modelling approach. *Journal of Refugee Studies*, *32*(Special Issue 1), i172–i182.

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the international AAAI conference on web and social media*, *8*(1), 216–225.

Inuwa-Dutse, I., Liptrott, M., & Korkontzelos, I. (2020). Migration and refugee crisis: a critical analysis of online public perception. *arXiv preprint arXiv:2007.09834*.

Kaufmann, E. (2019). Can narratives of white identity reduce opposition to immigration and support for hard Brexit? a survey experiment. *Political Studies*, *67*(1), 31–46.

Kreis, R. (2017). #refugeesnotwelcome: Anti-refugee discourse on Twitter. *Discourse & Communication*, *11*(5), 498–514.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.

Maggini, N., & Chiaramonte, A. (2019). Euroscepticism behind the victory of Eurosceptic parties in the 2018 Italian general election: Not exactly. *Journal of Common Market Studies*, *57*, 77–89.

Meltzer, C. E., Eberl, J.-M., Theorin, N., Heidenreich, T., Strömbäck, J., Boomgaarden, H. G., & Schemer, C. (2021). Media effects on policy preferences toward free movement: evidence from five EU member states. *Journal of Ethnic and Migration Studies*, *47*(15), 3390–3408.

Menshikova, A., & Van Tubergen, F. (2022). What drives anti-immigrant sentiments online? A novel approach using Twitter. *European Sociological Review*.

Mohammad, S. (2016). A practical guide to sentiment annotation: Challenges and solutions. In *Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis* (pp. 174–179).

NOS Nieuws. (2022). *Asielinstroom beperken of niet? De verschillen tussen enquêtes verklaard.* (https://nos.nl/artikel/2445270-asielinstroom-beperken-of-niet-de-verschillen-tussen-enquetes-verklaard. Retrieved on: 2022-11-23.)

OECD. (2015). *Migration Policy Debates – No.7, 2015* (No. 7). (https://www.oecd.org/migration/Is-this-refugee-crisis-different.pdf. Retrieved on: 2022-12-08.)

Papademetriou, D. G., & Banulescu-Bogdan, N. (2016). Understanding and addressing public anxiety about immigration. *Transatlantic Council on Migration. Migration Policy Institute*.

Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A primer in BERTology: What we know about how BERT works. *Transactions of the Association for Computational Linguistics*, *8*, 842–866.

Rowe, F., Mahony, M., Graells-Garrido, E., Rango, M., & Sievers, N. (2021). Using Twitter data to monitor immigration sentiment. In *Practitioners' guidebook. international organization for migration.* United Nations. Retrieved from https://github.com/fcorowe/twitter_migatittudes

Siapera, E., Boudourides, M., Lenis, S., & Suiter, J. (2018). Refugees and network publics on Twitter: Networked framing, affect, and capture. *Social Media+ Society*, *4*(1), 1–21.

Simon, R. J., & Lynch, J. P. (1999). A comparative assessment of public opinion toward immigrants and immigration policies. *International migration review*, *33*(2), 455–467.

Sniderman, P. M., Hagendoorn, L., & Prior, M. (2004). Predisposing factors and situational triggers: Exclusionary reactions to immigrant minorities. *American political science review*, *98*(1), 35–49.

Theorin, N., & Strömbäck, J. (2020). Some media matter more than others: Investigating media effects on attitudes toward and perceptions of immigration in Sweden. *International Migration Review*, *54*(4), 1238–1264.

Van der Schelde, A. (2022). *Meeste Nederlanders zien fatsoenlijke opvang asielzoekers als morele plicht.* (https://www.ioresearch.nl/actueel/meeste-nederlanders-zien-fatsoenlijke-opvang-asielzoekers-als-morele-plicht/. Retrieved on: 2022-11-23.)

Van Klingeren, M., Boomgaarden, H. G., Vliegenthart, R., & De Vreese, C. H. (2015). Real world is not enough: The media as an additional source of negative attitudes toward immigration, comparing Denmark and the Netherlands. *European Sociological Review*, *31*(3), 268–283.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wilkes, R., Guppy, N., & Farris, L. (2008). "No thanks, we're full": Individual characteristics, national context, and changing attitudes toward immigration. *International Migration Review*, *42*(2), 302–329.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... others (2019). Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.

Zimmermann, K. F., Bauer, T. K., & Lofstrom, M. (2000). Immigration policy, assimilation of immigrants and natives' sentiments towards immigrants: Evidence from 12 OECD-countries. *IZA Institute of Labor Economics Discussion Paper Series*.

APPENDIX A: HAND-ANNOTATION GUIDELINES

The interpretation of sentiment information in text is highly subjective, which can lead to disparity in annotations depending on the human judge. To ensure consistency and high-quality annotations in the evaluation set, the author adhered to a number of annotation guidelines inspired by the *simple* and *semantic role-based* sentiment frameworks proposed in Mohammad (2016). Tweets are primarily evaluated by their use of sentiment-laden language, and by the emotional state that is expressed or can be inferred from the message. For example, if profanity or derogatory language is used, the tweet would be labelled 'negative' regardless of the target of the opinion. Accordingly, tweets that convey a sense of fostering or empathy are marked positive, to the extent that the text contains explicit or implicit clues about the compassionate emotional state of the speaker. In any case, the personal views of the annotator about the message or target of opinion are to be disregarded.

A difficult yet frequent case in the news tweet samples is the occurrence of neutral tweets that report valenced events, e.g. violence and harassment by or towards migrants, death or rescue at sea, or asylum seekers forcibly sleeping outside versus the opening of a new shelter. In such cases, special consideration is given to the framing of the situation described, and the emotions that the tweet is expressing or evoking in a reader. An illustrative example is provided by the news tweets in Table 4 that refer to the same event, namely a novel proposal of the British government to transfer irregular migrants to Rwanda, but the second one contains additional sentiment-laden words, i.e. 'shock' (*shock*) and 'biglietto di sola andata' (*one-way ticket*):

Table 4: Italian news tweets reporting valenced events.

Tweet	**Label**
Il piano di Johnson sui migranti: «I richiedenti asilo entrati illegalmente andranno in Ruanda»	neutral
Annuncio ***shock*** *del primo ministro britannico Johnson sull'immigrazione. I richiedenti asilo arrivati nel Regno Unito senza un regolare permesso d'ingresso saranno trasferiti in Ruanda* ***con un biglietto di sola andata****. In cambio, il Paese africano riceverà 145 milioni di euro.*	negative

The same applies to tweets that describe one side benefiting over another, without the text itself using positive or negative language (e.g. the tweet *'VVD party supports law regarding distribution of asylum seekers'* would be considered neutral). By contrast, usage of words such as 'full to the brim', 'weak legal basis' or 'millions and millions of refugees' are considered to contain sentiment beyond the mere delivery of the factual contents, and are thus valenced. In short, if the speaker can be presumed to be in a negative or positive emotional state while posting, or to have a negative or positive attitude towards the target of opinion, the tweet is marked accordingly. However, if the emotional state of the speaker is unknown and no primary target of opinion can be identified, the tweet can be considered neutral. In such cases, the nature of the event and the dominant tone or frame of the message are considered: if the sentimental impact of the tweet on most readers would be negative (positive), the tweet is also considered negative (positive). This is illustrated by the news tweets in Table 5 that report on asylum seekers sleeping outdoors, with the second one mentioning 'pijn' (*pain*) and 'verschrikkelijke kou' (*horrible cold*):

Table 5: Dutch news tweets with their respective dominant tones.

Tweet	**Label**
Enkele honderden asielzoekers Ter Apel moeten buiten slapen	neutral
Buitenslapers Ter Apel hebben ***pijn*** *door* ***'verschrikkelijke kou'*** *vannacht*	negative

In the reply tweets, the most challenging examples are messages that express different sentiment towards different targets of opinion (e.g. *'For regular migrants, yes. For illegals, prison and deportation.'*). These are generally labelled neutral, as no categories for 'both', 'mixed' or 'neither positive or negative' exist in either of the machine learning models used for this analysis. If an event or decision is merely questioned, rhetorically or not, the tweet is also considered neutral. However, in line with the aforementioned guidelines, tweets that contain clues about the speaker's emotional state with respect to the target of opinion, such as empathy, sarcasm, ridicule, or suggestive terminology or emojis, could be considered valenced (e.g. *'So what?'* and *'Everything for the show'* are inferred to be negative attitudes).

Evaluation Set

Table 6 shows the distribution of human-annotated labels in the evaluation set. The full sets of annotated tweets are available in the author's repository on Github: `https://github.com/melaniewaterham/thesis`.

Table 6: Distribution of human-annotated tweets in the evaluation set.

		Label		
		negative	neutral	positive
News	Dutch ($n = 50$)	23	22	5
	Italian ($n = 50$)	23	19	8
	Total News	**46**	**41**	**13**
Replies	Dutch ($n = 50$)	31	17	2
	Italian ($n = 50$)	29	16	5
	Total Replies	**60**	**33**	**7**
	Total ($n = 200$)	**106**	**74**	**20**

APPENDIX B: DISTRIBUTION OF SENTIMENT BY OUTLET (VADER)

This Section contains Figures that relate to sub-question *Q.3* regarding the difference in sentiment by outlet according to VADER, as addressed in Section 4.3.

Figure 12: Distribution of VADER compound scores by outlet in news_nl.

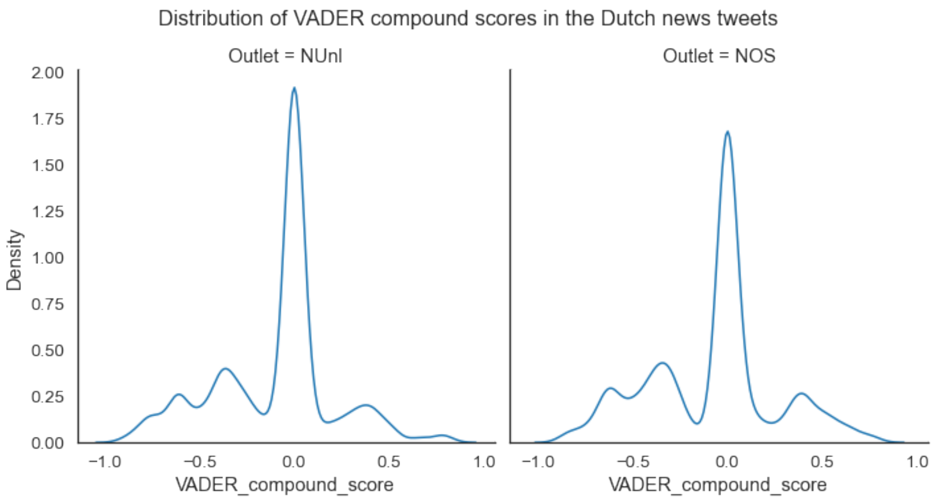


Figure 13: Distribution of VADER compound scores by outlet in replies_nl.

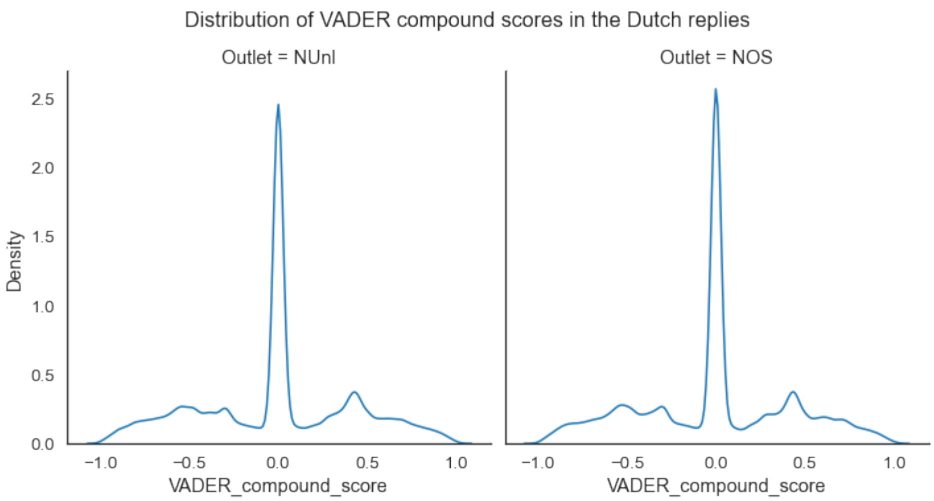


Figure 14: Distribution of VADER compound scores by outlet in news_it.

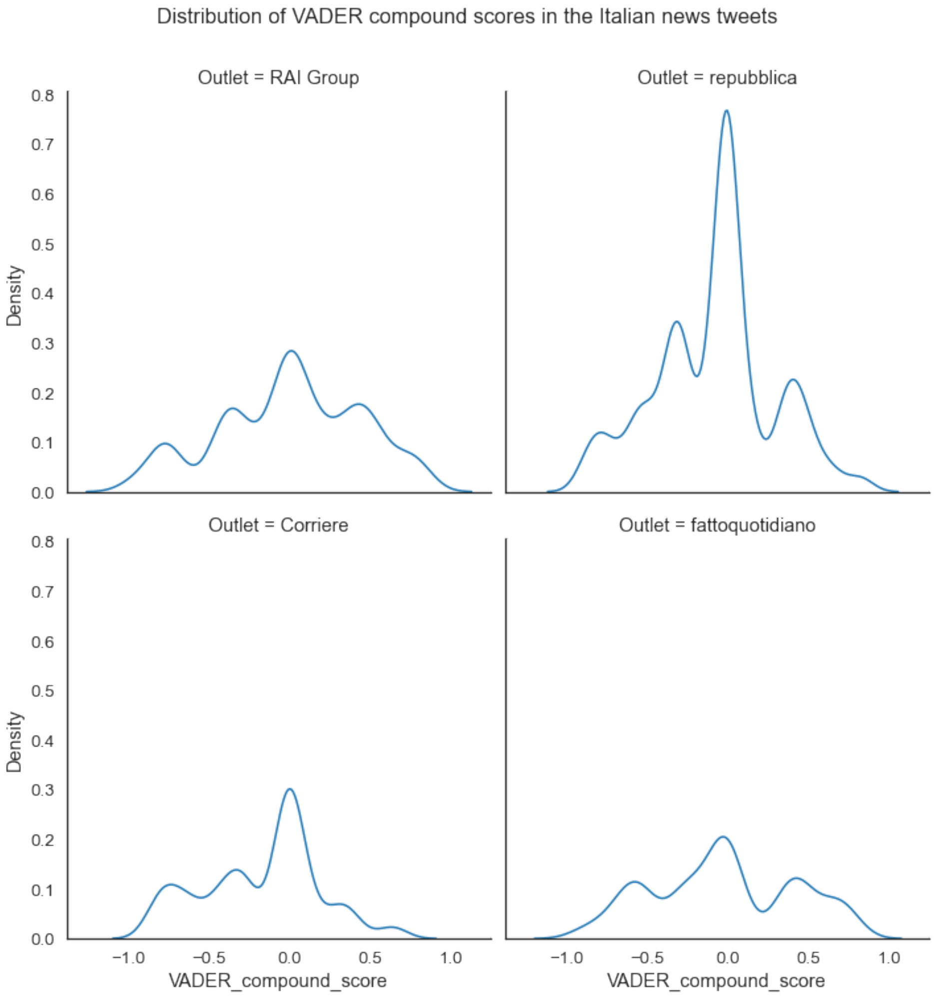
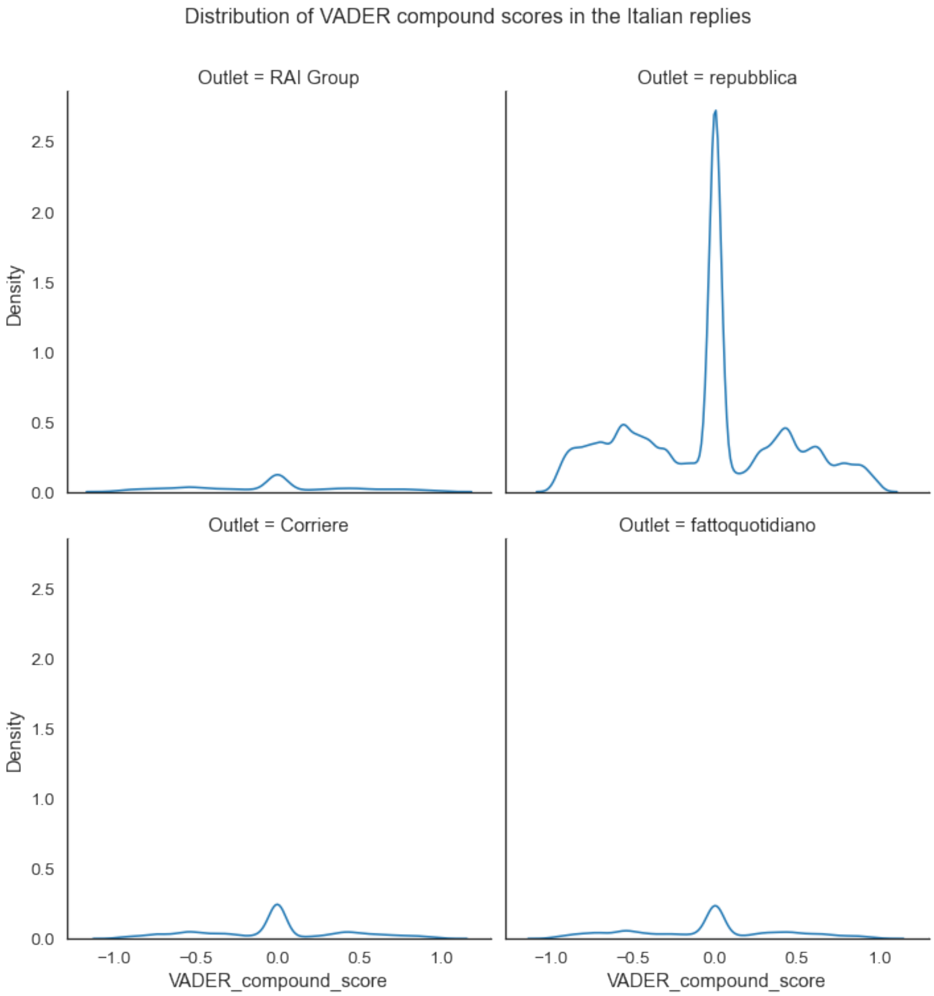


Figure 15: Distribution of VADER compound scores by outlet in replies_it.



APPENDIX C: TOP FIVE NEWS TWEETS BY INTERACTION

The Tables below present the news tweets that provoked the highest number of replies by country, as addressed in Section 4.4.

Table 7: Top five Dutch news tweets by reply count.

Tweet	Reply Count
Adviesraad: Geef vluchtelingen met verblijfstatus voorrang bij huurwoningen	1166
Kabinet wil asielcrisis oplossen met 20.000 huizen en meer geld	967
Buitenslapers Ter Apel hebben pijn door 'verschrikkelijke kou' vannacht	837
Zo smerig zijn de toiletten in Ter Apel	776
300 asielzoekers in Ter Apel sliepen vannacht buiten, hoogste aantal tot nu toe	709

Table 8: Top five Italian news tweets by reply count.

Tweet	Reply Count
Insulti razzisti, pugni e calci contro un migrante accusato di essersi spogliato in pubblico	1807
Decreto rave, Elly Schlein: "Piantedosi ritiri la norma". E sui migranti: "La battaglia non si fa bloccando le persone in mare".	990
Primo schiaffo dell'Ue a Meloni. E a Catania i migranti a bordo sbarcano.	889
No a gay, migranti e aborto: le crociate dell'ultracattolico che sfilava con i neonazisti (scopri la app) #RepSelezione	837
Soumahoro: "Stipendi più alti e cittadinanza per i figli, i migranti non sono schiavi"	492