Using State-of-the-art Emotion Detection Models in a Crisis Communication Context

de Bruyne, L.; van der Meer, T.G.L.A.; de Clercq, O.; Hoste, V.

DOI
10.5117/CCR2024.1.1.DEBR

Publication date
2024

Document Version
Final published version

Published in
Computational Communication Research

Amsterdam University Press

COMPUTATIONAL COMMUNICATION RESEARCH 6.1 (2024) 1–45
HTTPS://DOI.ORG/10.5117/CCR2024.1.1.DEBR

Using State-of-the-art Emotion Detection Models in a Crisis Communication Context

Luna De Bruyne
Computational Linguistics and Psycholinguistics Research Group (CLiPS), University of Antwerp;
LT³, Language and Translation Technology Team, Ghent University

Toni G.L.A. van der Meer
Amsterdam School of Communication Research (ASCoR), University of Amsterdam

Orphée De Clercq
LT³, Language and Translation Technology Team, Ghent University

Véronique Hoste
LT³, Language and Translation Technology Team, Ghent University

Abstract

Times of crisis are usually associated with highly emotional experiences, which often result in emotionally charged communication. This is especially the case on social media. Identifying the emotional climate on social media is imperative in the context of crisis communication, e.g., in view of shaping crisis response strategies. However, the sheer volume of social media data often makes manual oversight impossible. In this paper, we therefore investigate how automatic methods for emotion detection can aid research on crisis communication and social media. Concretely, we investigate two Dutch emotion detection models (a transformer model and a classical machine learning model based on dictionaries) and apply them to Dutch tweets about four different crisis cases. First, we perform a validation study to assess the performance of these models in the domain of crisis-related tweets. Secondly, we propose a framework for monitoring the emotional climate on social media, and assess whether emotion detection models can be used to address the steps in the framework.

Keywords: sentiment analysis, emotion detection, automated content analysis, evaluation, validity, crisis communication

Introduction

Understanding the role of affect in communication has been an important research direction in communication science. Extant research has been performed on the affect that is conveyed in news media, political communication or messages from organizations and its effect thereof on various actors (e.g., Kim, 2015; Rhodes & Vayo, 2019; Schoofs & Claeys, 2021).

Especially in the context of crisis communication, affect plays an important role. Times of crisis are usually associated with highly emotional experiences, which often result in emotionally charged communication. Analyzing these emotions might inform crisis managers from different organizations, ranging from public-health organizations to for-profit corporations, for shaping communication strategies, and can inform them about the state of the crisis (Schultz et al., 2012). Panic, for example, might indicate a potential escalation of the crisis, while more positive emotions might hint toward a solution of the crisis.

In the situational crisis communication theory (SCCT) by Coombs (2007), one of the central theories in the field of crisis communication, emotions are allocated a pivotal role. Rooted in attribution theory, SCCT highlights sympathy and anger as the relevant emotions, which basically comes down to a positive-negative dichotomy. However, more fine-grained emotional information is often desired to obtain a more detailed understanding of the communication. When dealing with crises, the distinction between anger, sadness or fear, although all negative emotions, might be imperative to accurately monitor the crisis evolvement or identify the most relevant response strategy.

In addition to its emotional dimension, crisis communication often comes with an overload of information, especially in this day and age where many of our interactions take place online. To address the reality that the public is increasingly turning to social media during crises, B.-F. Liu et al. (2012) introduced the social-mediated crisis communication model (SMCC). This model aids crisis managers in comprehending how the public generates, consumes, and disseminates crisis information through social media and other channels. In this model as well, the importance of dealing with negative emotions uttered on social media is emphasized.

The sheer volume of social media data, especially of the overload of communication in times of crisis, often makes manual oversight impossible, necessitating the use of automated techniques. Automated content analysis by means of computational methods allow for a time and cost-efficient solution for analyzing digital data at a greater scale. In the context of crisis communication, van der Meer (2016) already advocated for the use of such techniques, e.g., to determine whether the tone in a text is positive, negative or neutral (sentiment analysis), to get insights in the central actors of a crisis or to automatically code frames. We believe that using computational techniques to analyze more fine-grained emotional information could also be a desired method in social media-focused crisis communication and management. In this context, we propose to use the term 'emotional climate monitoring', which refers to the the fact that we are interested in general trends regarding the public's conveyed emotions, rather than aiming to accurately analyze the emotions in a specific document.

Sentiment analysis tools have already been adopted and validated in (crisis) communication research (Boukes et al., 2020; Van Atteveldt et al., 2021), but this is much less the case for the fine-grained analysis of emotions, which is referred to as automatic emotion recognition or emotion detection in the field of Natural Language Processing (NLP). In this paper, we therefore want to investigate the potential of automatic emotion detection models in the context of crisis communication and social media. The main research question of this paper is thus the following:

> **RQ:** *How can emotion detection models aid research on crisis communication and social media?*

Although advances in NLP have led to sophisticated emotion detection models based on machine learning, the task of categorizing emotions in text remains a challenging task, even for humans. Therefore, it is necessary to first validate existing models, and assess whether automatic emotion detection, given the challenging nature of the task, is even worth pursuing in the context of crisis communication. Emotion detection models have been assessed by means of intrinsic evaluation in NLP studies, which in practice means that the models are tested on a part of the dataset on which the model was trained. Especially transformer models have been proven powerful for the task of emotion detection and were fine-tuned and evaluated on various datasets, including reviews (De Geyndt et al., 2022), tweets (Barbieri et al., 2020) and conversations (Zhu et al., 2021). However, we will take on a somewhat different approach and will validate the output of emotion detection models – that were not specifically developed for usage in a crisis communication context, nor fine-tuned on this domain, but merely used in inference mode – on unlabeled, crisis-related data. This thus leads to a first sub-question that deals with **validation**:

> **SQ1:** *Does emotion detection perform well enough to be worth*

pursuing in the context of crisis communication, and if so, which model should be used?

After looking into this validation question, we will assess the application potential of emotion detection models in the context of social media and crisis communication practice and research. To this end, we propose a framework for emotional climate monitoring on social media, consisting of three steps: analyzing the temporal evolution of emotions communicated by social media users in the context of a certain crisis; identifying key topics that are associated with the found emotions; and analyzing the influence of emotional content on social media traffic. The proposed framework is visualized in Figure 1. We will thus assess the application potential of the emotion detection model based on the steps proposed in this framework. This results in a second sub-question that involves **application testing**:

SQ2: *Can emotion detection models be applied for mapping the temporal evolution of emotions, identifying associated topics, and predicting online traffic?*

To this purpose, we will focus on the social media platform X (formerly known as Twitter), from which we collect Dutch tweets related to four recent crises in the Netherlands. In order to select a variety of crisis types, each case represents one of the four areas of crisis communication as distinguished by Coombs and Holladay (2022): (i) political crises, (ii) natural disasters, (iii) public health crises, and (iv) organizational crises. The selected cases are

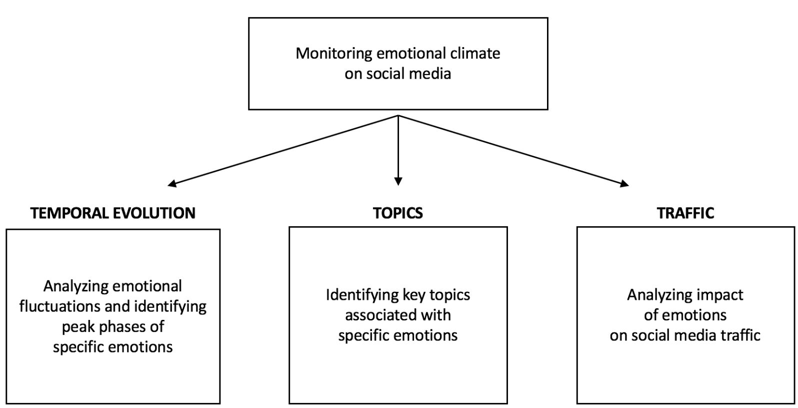


Figure 1: Steps in social media emotional climate monitoring.

the childcare benefits scandal as political crisis, the exceptional floods from Summer 2021 as natural disaster, the COVID-19 pandemic as health crisis and the controversies about the talent show The Voice of Holland in the beginning of 2022 as organizational crisis.

Two Dutch emotion detection models will be tested. We only focus on machine learning methods, as these have been found to reach better performance than rule-based or lexicon-based methods in both NLP studies and validation studies on sentiment analysis (B. Liu, 2020; Van Atteveldt et al., 2021). The machine learning methods investigated in this paper are a state-of-the-art transformer model that was developed in the EmotioNL[1] project (De Bruyne et al., 2021b) and a more traditional machine learning system (support vector machine) with opinion lexicons as features.

The remainder of this paper is structured as follows: first, we discuss background information on emotions in crisis communication research and summarize different approaches to the automated measurement of emotions. We then give a description of the resources and methods, including an overview of the emotion detection models and data, used in this paper. The subsequent section reports the results of our validation and application testing studies. We end this paper with a conclusion.

Background

Emotions in crisis communication

One of the most prominent crisis communication frameworks is the Situational Crisis Communication Theory (SCCT) developed by Coombs (2007). This theory offers guidelines for constructing response strategies that maximize the reputational protection of organizations in times of crisis. In the original presentation of SCCT, emotions are (parallel with organization reputation) presented as the link between crisis responsibility and the behavioral intentions of stakeholders (see Figure 2). Crisis response strategies are thus used to reduce negative affect and to prevent negative behavioral intentions (e.g., reduced purchase intentions or negative word of mouth). As SCCT is rooted in attribution theory, anger and sympathy are the core emotions in this framework: the presence of crisis responsibility triggers anger and is associated with reputational threat, while absence of responsibility (e.g., when the crisis results from situational factors) is more often associated with sympathy.

Choi and Lin (2009) proposed to explore emotions other than anger and

[1] https://research.flw.ugent.be/en/projects/emotionl-emotion-detection-dutch

sympathy and investigated how they influenced the dynamics of SCCT. Inspired by the work of Weiner (1986), they distinguished attribution-dependent and attribution-independent emotions, where the latter category refers to general emotional reactions to events and the former category refers to emotions that involve attribution processes and are thus related to the cause of events. They identified 11 crisis emotions, of which anger, fear, surprise, worry, contempt and relief were found to be associated with crisis responsibility and were thus considered attribution-dependent emotions, while alert and confusion were not found to be associated with crisis responsibility and were considered attribution-independent emotions. A main finding was that attribution-dependent emotions like anger coexist with attribution-independent emotions like alert and confusion. This, together with the finding that different emotions are associated with different behavioral intentions (e.g., boycotting was associated with anger but not with fear), emphasizes the need to take into account more (and more fine-grained) emotions when studying or practicing crisis communication.

Jin (2010) as well explicitly emphasized the need to examine the general public's emotional state in a crisis context. This emphasis extends to understanding how emotional factors impact the public's coping mechanisms and their responses to the used communication strategies and to the crisis itself.

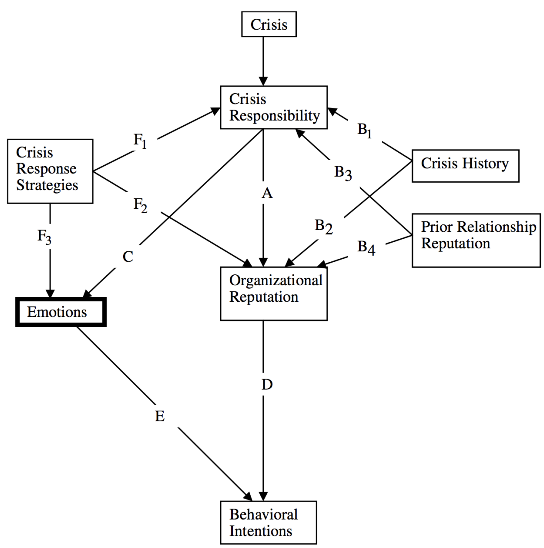


Figure 2: Emotions in the crisis situation model of SCCT, adapted from Coombs (2007).

She stated that, during public health crises or natural disasters, monitoring emotions is crucial to spot levels of outcry, examine whether there is support for government measures and even to track the evolvement of the crisis.

These ideas were embodied in the Integrated Crisis Mapping (ICM) model (Jin et al., 2010). The model maps different crises with their dominant and secondary emotions onto two dimensions: the primary publics' coping strategy (from conative to cognitive coping) and the organizational engagement with the crisis (ranging from low to high). The goal of ICM is to provide a framework for shaping crisis responses from an emotion-based perspective, rather than from a situation-based angle like in SCCT.

Later, Jin, Liu, and Austin (2014) explored the occurrence of various negative emotions in the context of the social-mediated crisis communication model (SMCC). This theoretical framework examines the relationship between an organization in crisis and key publics through social media, traditional media and offline word-of-mouth communication (Austin et al., 2012; B.-F. Liu et al., 2012). By taking into account the crisis origin and the form and source of crisis information, the SMCC model assists crisis managers in determining whether and how to engage with influential online content creators. Jin, Liu, and Austin (2014) therefore investigated what crisis emotions publics are likely to feel when exposed to crisis information, taking the crisis origin into account. Slightly deviating from Choi and Lin (2009), they found three clusters of emotions, one indeed being attribution-independent emotions (e.g., fear) but the attribution-dependent emotions being split up into external (e.g., anger) and internal (e.g., shame) attribution-dependent emotions. In further work, Jin, Liu, Anagondahalli, and Austin (2014) made use of these three emotion clusters to develop a scale for measuring publics' crisis emotions, in order to help crisis communication professionals in crafting crisis responses.

Other researchers as well have advocated for aligning crisis response strategies with the emotions of stakeholders. Lu and Huang (2018), for example, propose an emotion-cognition dual-factor model of crisis communication, which analyzes how emotions influence the cognitive processing of crisis information and how this affects their attitudes and behaviors toward organizations.

Echoing ideas from SCCT, SMCC as well as ICM, Vignal Lambret and Barki (2018) furthermore argue that social media require adapted crisis response strategies which are more stakeholder-centric (instead of merely organization-centric) and where emotions should be taken into account. The framework they propose is therefore based on the relation between the origin of the crisis, the degree of crisis responsibility, and the stakeholders' emotions in reaction to the crisis. The emotions of interest in this framework are sympathy, sadness, fright and anger.

Marx et al. (2021) as well posit that response strategies might be more effective if they are aligned with what they call "the prevalent emotional climate in stakeholder communication". They particularly highlight the role of social media and present a social media analytics approach to monitor (negative) emotions toward organizations in crisis. The authors performed a case study on Tweets about the airplane manufacturer Boeing, which suffered from a corporate crisis in 2019 due to repeated fatal crashes of their Boeing 737 MAX model. They analyzed the development of emotions over time and were able to identify peak phases in the temporal evolution and link those to specific events in the crisis. Again, these findings suggest that publics' emotions can be used to inform response strategies.

Methods for measuring emotions

In studies where stakeholders' feelings (or their attitude toward response strategies) in crisis situations are analyzed, emotions are mostly measured by means of manual coding (e.g., Choi & Lin, 2009; Jin et al., 2010; Vignal Lambret & Barki, 2018) or by conducting surveys (e.g., Jin, Liu, & Austin, 2014). However, audiences are increasingly turning to social media during crises (Austin et al., 2012), resulting in an enormous amount of communicative data of which the manual analysis would be heavily time and cost intensive.

Automated content analysis by means of computational methods allows for a time and cost-efficient solution for analyzing digital data at a greater scale. Moreover, as van Atteveldt and Peng (2018) pointed out, computational methods create new opportunities in communication science research compared to traditional approaches. Namely, the opportunity to a) study real behavior rather than relying on self-reported data, b) analyze actual social environments instead of simulations in lab settings, c) study more subtle relations or effects in subpopulations, by leveraging more complex models which these larger datasets allow for, and d) make collaborative research possible because digital data and resources can be easily shared. These advantages, together with the emergence of powerful and cheap processing power and computing infrastructure over the last decades, has fueled the development and usage of computational methods in communication science.

An additional advantage, more specific to the analysis of emotions, is that automated methods allow for a consistent and more objective judgment, whereas manual coding often suffers from both inter- and intra-coder inconsistencies (De Bruyne et al., 2021a; Troiano et al., 2021). Computational models, on the other hand, do make consistent decisions, even though the accuracy of automated approaches is not perfect. Note however, that performance of automated emotion detection has improved over the years, as their development follows general advancements in NLP (starting from off-the-shelf dictionary-based approaches to classical machine learning methods, deep learning and at last state-of-the-art transformer models).

For sentiment and emotion analysis, many off-the-shelf methods are available, mostly relying on dictionaries. Such dictionaries or sentiment lexicons consist of long lists of (opinion) words that have been assigned an associated emotion label or score. After some pre-processing steps (like lowercasing, tokenization and lemmatization), string matching is performed to find the opinion-bearing words in a target text, and by looking up all the corresponding emotion values, the overall sentiment or emotion values of that text can be determined automatically. Popular lexicon-based sentiment analysis methods used in communication science are for example the Linguistic Inquiry and Word Count or LIWC (Pennebaker et al., 2007), SentiStrength (Thelwall et al., 2010) and the General Inquirer (Stone et al., 1966).

The performance of dictionary-based approaches is often disappointing, as lexical coverage is a big problem (B. Liu, 2020). Such systems are only as good as the words that appear in the lexicon, whereas language is very productive (think for example of the covid-related surge of neologisms or new usages of existing words and phrases like ‘covidiot’, ‘self-quarantine’ or ‘flatten the curve’). Moreover, sentiment and emotion are highly dependent on the domain. In fact, B. Liu (2020) mentions four main issues with opinion lexicons: a) a word can have a different sentiment polarity depending on the application domain or sentence context, b) a sentence comprising opinion words does not necessarily express sentiment (e.g. questions or conditional sentences), c) sarcasm, which for example often occurs in political opinions, causes an opposite polarity, and d) sentiment can be expressed implicitly, i.e. without using explicit opinion words.

However, dictionary-based approaches are still widely used in communication research (Jonkman et al., 2020; Marx et al., 2021; Schultz et al., 2012), as they also have some important advantages: first of all, they do not need training data, which makes them much easier and less (computationally) costly to use. Secondly, lexicons can be crafted by domain experts. The Linguistic Inquiry and Word Count (LIWC) for example, which analyzes psychological properties in word use by means of dictionaries and includes a sublexicon for emotion, was crafted by psychologists (Pennebaker et al., 2007), but also communication scientists have crafted lexicons that can be used to measure sentiment and emotion (e.g., Risius & Akolk, 2015; Soroka, 2006).

In the field of Computational Linguistics and NLP, much effort has been devoted to employing machine learning for the task of sentiment analysis. For this task supervised machine learning techniques are used which rely on so-called training data, which is coded (manually or automatically derived, e.g., by using hashtags on Twitter) with sentiment or emotion labels. Traditional machine learning approaches rely on feature-based models, which means that features are being engineered capturing the characteristics of the textual data in numerical values. For sentiment analysis a common way of transforming the text data to such features is the bag-of-words model, which extracts word frequencies (or n-gram frequencies, where *n-grams* are *n* consecutive words) from the texts. Also sentiment values from dictionaries have often been used as features. Next, machine learning algorithms (e.g., Naive Bayes, Support Vector Machine or Logistic Regression) learn patterns from these feature representations which are linked to the output labels. The learned model can then be used to predict the sentiment label for unseen texts by deriving the same features.

Over the years, the approaches for representing text as features and the classification algorithms became more and more sophisticated by the rise of deep learning. Compared to the previous approaches, in deep learning the feature extraction part has been automated, in that the text is represented as so-called word embeddings with high dimensionalities. These word embeddings are often pre-trained on other, larger datasets, which resolves the vocabulary coverage problem. Word embeddings form the input of a multi-layered neural network, which learns to make predictions by processes of gradient descent and backpropagation. When dealing with text, the most common neural networks are recurrent neural networks and more specifically LSTM (*Long Short-Term Memory*) networks, which can handle sequential data and thus can take into account word order.

More recently, transformer models are considered the state of the art in NLP. Tranformers are neural networks that learn context by tracking relationships in sequential data (e.g., words in a sentence). These models are pre-trained on huge amounts of data. An interesting aspect of these (large) language models (LLMs), such as the well-known BERT (Devlin et al., 2019) or GPT (Radford et al., 2019), is that they are very successful at transfer learning. This means that these models can be fine-tuned on a specific target task on just a limited amount of labeled data. They achieve state-of-the-art performance for many NLP tasks, including sentiment analysis (Devlin et al., 2019).

For analysing more fine-grained emotions instead of just sentiment polarity, the same tendencies have taken place in the field of NLP. Emotion analysis was first performed by means of dictionaries and linguistic rules (Chaumartin, 2007) or traditional machine learning (Chaffar & Inkpen, 2011; S. Mohammad, 2012), followed by deep learning (Jabreel & Moreno, 2019; Schuff et al., 2017) and transformers (Acheampong et al., 2021). For emotion detection in Dutch, most recent advances were obtained in the EmotioNL project (De Bruyne et al., 2021b), which used the RobBERT transformer model (Delobelle et al., 2020). This model is the current state of the art for Dutch, and will therefore be validated in this paper together with a more classical machine learning approach with dictionary values and n-grams as features.

Resources and Methods

In this section, we elaborate on the methods and resources used in this paper. First, we will describe the used emotion detection models. We only focus on machine learning methods, as these have been found to reach better performance than rule-based or lexicon-based methods in NLP studies as well as in validation studies on sentiment analysis in communication science (B. Liu, 2020; Van Atteveldt et al., 2021). We are mainly interested in the transformer-based method, as this is considered the best approach currently available. Moreover, as a Dutch emotion detection model was developed in the EmotioNL project (De Bruyne et al., 2021b) and released on Huggingface[2], it is possible to use this model in an almost off-the-shelf manner, meaning that we will not further fine-tune the model on the crisis communication data but merely use it in inference mode. We want to investigate whether a more traditional machine learning approach, which requires less computational resources, can compete with the transformer-based method. Here as well, we will use a model trained on a background corpus (the same one that was used in the EmotioNL project), without specifically training on crisis-related data. This approach best reflects how communication professionals would incorporate emotion detection models in a real-world setting. In this section, we will first describe the transformer-based and classical machine learning model. Then, the data that are used for the validation and

[2]https://huggingface.co/lunadebruyne/emotionl-classification

application testing studies is presented. Finally, we give a description of the validation and application testing experiments that are conducted in this paper.

Emotion detection models

Transformer model: EmotioNL

The Dutch transformer model EmotioNL (De Bruyne et al., 2021b) is a fine-tuned version of the pre-trained RobBERT model (Delobelle et al., 2020). The dataset used for fine-tuning and evaluating EmotioNL consisted of two subsets with respectively 1,000 tweets and 1,000 captions from reality TV-shows.[3] Each instance in both subsets was labeled with one out of the six categories anger, fear, joy, love, sadness or neutral. Models were fine-tuned on each of the subsets separately, and on the combination of the two. The authors released their original code on Github[4] (which we used to replicate the combined model and use it in the current study), and recently, the authors also made their models available on Huggingface.

In the original paper, the models were evaluated on the EmotioNL subsets using 10-fold cross validation. The combined model obtained an accuracy of .515 on the tweets subset and .517 on captions, or a macro-averaged F1 of .381 and .396 on the respective domains (De Bruyne et al., 2021b). The authors also calculated an additional metric, which they called cost-corrected accuracy and which allows for a fairer evaluation of emotion detection models. The metric takes into account the severity of a misclassification, by introducing misclassification weights for each class pair (De Bruyne et al., 2021b). Misclassifications for class pairs within the same polarity (e.g., misclassifying fear as anger) receive a lower cost weight than class pairs with an opposite polarity (e.g., misclassifying love as sadness).[5] Using this metric, the authors reported a score of .670 for tweets and .677 for captions.

These scores are relatively low, but in line with results from other papers for English emotion detection. In the benchmarking paper by Barbieri et al. (2020), for example, a RoBERTa model was used for emotion detection in tweets (with classes anger, sadness, joy and optimism) for which their highest F1-score was .316. We see several reasons for these low scores. Firstly,

[3]The annotated dataset was made available at https://lt3.ugent.be/resources/emotionl/.

[4]https://github.com/LunaDeBruyne/Mixing-Matching-Emotion-Frameworks

[5]Misclassifications with an opposite polarity (e.g., love-sadness) receive a cost weight of 1, misclassifications involving the neutral category (e.g, joy-neutral) receive a cost weight of 2/3 , misclassifications within the same polarity (e.g., fear-anger) receive a cost weight of 1/3. Correct classifications have a cost of 0.

tweets are short and often lack context, they often contain implicit emotions (contrary to reviews, in which the emotion expression is more explicit), and finally, determining emotions from texts is even difficult for humans, especially when differentiating between emotion categories within a same polarity (e.g., distinguishing sadness from fear and anger, or joy from love).

Classical machine learning model with dictionaries

The transformer model is compared to a classical machine learning model, namely a support vector machine with dictionary values and n-grams as features. To allow for comparison with the transformer model, we use the publicly available EmotioNL dataset to train an SVM that classifies texts as either anger, fear, joy, love, sadness or neutral.

Eight open-source Dutch sentiment and emotion dictionaries are used, namely Pattern (De Smedt & Daelemans, 2012), Duoman (Jijkoun & Hofmann, 2009), LIWC (Boot et al., 2017), NRC Emotion (S. M. Mohammad & Turney, 2013), NRC VAD (S. M. Mohammad, 2018), Memolon (Buechel et al., 2020), BabelSenticnet (Vilares et al., 2018) and the VAD norms by Moors et al. (2013). The tweets from the EmotioNL dataset are lowercased, tokenised and lemmatised. Then, lexicon values are obtained through a lookup in each affect lexicon, and the values are averaged over the words in the target sentence. These features are complemented with word n-grams ($n = [1, 2]$, i.e. word frequencies and frequencies for 2 consecutive words), and character n-grams ($n = [3, 4, 5]$, i.e. frequencies for 3, 4 or 5 consecutive characters).

To give a first comparison with the transformer-based approach, we provide an intrinsic evaluation by performing 10-fold cross validation and calculating classification accuracy, macro F1 and cost-corrected accuracy on the tweets and captions subsets. Using the default hyperparameters of SVM in the scikit-learn Python library,[6] the model obtained an accuracy of .415 and .428, an F1-score of .255 and .304, and cost-corrected accuracy of .563 and .580 for tweets and captions respectively. A summary of these metrics, compared to the metrics of the EmotioNL transformer model reported by De Bruyne et al. (2021b), is given in Table 1.

Note that we only evaluate one configuration of the model here: we did not experiment with different feature sets (although we made sure we had a wide range of relevant features, including most open-source lexicons available for Dutch), nor did we perform any parameter tuning. We intentionally

[6]Hyperparameters: regularization parameter C is 1.0; kernel type is 'rbf'; kernel coefficient gamma is 'scale'.

	Tweets			Captions		
Model	**F1**	**Acc.**	**Cc-Acc.**	**F1**	**Acc.**	**Cc-Acc.**
Transformer	0.381	0.515	0.670	0.396	0.517	0.677
SVM	0.255	0.415	0.563	0.304	0.428	0.580

Table 1: Intrinsic evaluation metrics of the transformer model and SVM model.

decided to not rely on heavy feature engineering or parameter tuning in order to minimize the risk of overfitting the model on the training data.

This intrinsic evaluation clearly shows that the transformer model outperforms the SVM. However, apart from performance in terms of accuracy and related metrics, another important criterion to take into account is computational cost. Transformer models generally need a GPU to be trained or fine-tuned, and even for inference (i.e., using the model to make predictions on unlabeled data, which is what we will be doing as a way of extrinsic evaluation in the remainder of this paper) a GPU is recommended (inference using a CPU is possible, but less efficient). Training and inference with SVMs, on the other hand, requires much less computational power. Therefore, we are interested in seeing the SVM's capabilities in an extrinsic validation setting as well. To this purpose we retrain the SVM on the full EmotioNL dataset.

Data

In the latest version of *The Handbook of Crisis Communication*, Coombs and Holladay (2022) distinguish four types of crises: political crises, public health crises, natural disasters, and organizational crises. In order to not restrict ourselves to one crisis type when investigating how emotion detection models can aid research on crisis communication and social media, but rather obtain general insights, we collect tweets about a recent crisis from each of those types. A description of the selected crisis cases is given in Table 2. As organizational crisis, we chose the scandal about The Voice of Holland which took place early 2022. We collected data from the period 01/01/2022 – 28/02/2022 with `#tvoh` as search query. As political crisis, we focused on the childcare benefits scandal, with `#toeslagenaffaire` (the original Dutch term to denote this crisis) as search query and time period 01/11/2020 – 31/07/2021. For the natural disaster, we collected tweets about the floods in the Netherlands and Belgium from Summer 2021 (search period

Organizational crisis: The Voice of Holland Dutch reality TV singing competition, broadcasted by RTL 4 and created by John De Mol for his production company Talpa. Candidates participate in a 'blind audition', in which they try to convince the jury of their singing talent, without the jury seeing the candidate. The jury consists of four coaches, who are known Dutch musicians. The twelfth season was suspended in response to sexual misconduct allegations, relating to band leader Jeroen Rietbergen, old coach Marco Borsato and coach Ali B. The allegations came to light following an investigation by journalist Tim Hofman for the YouTube program BOOS.	**Political crisis: Childcare benefits scandal (Dutch: *toeslagenaffaire*)** Dutch political scandal concerning false allegations of fraud regarding childcare benefits (childcare allowance) and the strict policy to pay back the allowances. The working procedure of the Tax and Customs Administration turned out to be unlawful and discriminatory (people with dual nationality, for example, were investigated more severely). The scandal has affected an estimated 26.000 parents and 70.000 children between 2004 and 2019. They had made – often minor – mistakes or were misled by fraudulent childminding agencies, and had to pay back the entire allowance.
Natural disaster: Floods Summer 2021 In mid-July 2021, the Belgian province of Luik, Belgian Limburg and the southern part of the Dutch province of Limburg were affected by exceptional floods as a result of persistent rain showers.	**Health crisis: COVID-19** At the end of 2019, the infectious disease COVID-19 broke out in Wuhan (China), which also reached Belgium and the Netherlands at the end of February 2020. These countries went into lockdown in mid-March.

Table 2: Description of crisis cases used in this study.

10/07/202 – 31/07/2021) with the flooding-related hashtags (`#watersnood OR #watersnoodLimburg OR #watersnoodramp OR #wateroverlast`) as search query. Finally, we chose the COVID-19 pandemic as health crisis, focusing on the beginning of the pandemic. We used (`#corona OR #coronavirus OR #coronacrisis OR #covid OR #covid19 OR #covid-19`) as search query in the time period 01/01/2020 – 31/08/2020. All search queries were provided with the language tag `lang:nl` in order to only pull Dutch data, and retweets were excluded using `-is:retweet`.

All data were crawled from the Twitter API by Twarc, a Python library for archiving Twitter JSON data, using an academic account. This resulted in a dataset of 18,502 posts about The Voice of Holland, 66,961 instances about the childcare benefits scandal, 9,923 instances related to the floods, and 609,206 COVID-19 related tweets (see Table 3). For every tweet, we have the following meta data: tweet ID, user ID, posting date and time, number of likes, retweets, replies and quotes.

Validation procedure

In order to test the validity of the emotion detection models in the context of crisis communication and social media (SQ1), we first applied the transformer-based model to the crisis-related tweets described above. For each of the four crisis cases, a subset was selected as validation set. To this purpose, we looked at the predictions of the transformer model and randomly picked 20 instances for every predicted label. This resulted in a set of 120 tweets (i.e., 20 of which were classified by the transformer as neutral, 20 as anger, 20 as fear, 20 as joy, 20 as love and 20 as sadness). The order of the tweets was randomized and the subset was then classified by the SVM as well.

Instead of evaluating the models by means of an intrinsic evaluation (like we did when describing the emotion detection models), we now go one step further and validate the output of the models in an extrinsic evaluation by judging its output directly by possible end-users. The output evaluation consists of assessing whether humans agree with the label predicted by the models, by letting them choose one out of four options: agree, not agree, doubtful agree, or impossible to judge.

The reason for this approach is twofold. Firstly, assessing the output of the systems allows for a fairer evaluation, as coders are often not sure about their own annotations and are forced to pick a label for obtaining a gold standard. By letting the coders assess the output, they can take into account their uncertainty (by picking 'maybe'), or indicate that it is impossible to judge the emotional content (e.g., due to a lack of context). Secondly, this approach is similar to real-world application settings, in which users of a tool can flag wrong predictions and suggest corrections on which the model can be retrained (Frasnelli et al., 2021).

We recruited three trained linguists who volunteered to judge the model predictions. They received four matrices to annotate (one for each crisis case), with the 120 tweets in the first column, and three labels in the following

Crisis case	Number of instances
The Voice of Holland	18,502
Childcare Benefits	66,961
Floods	9,923
COVID-19	609,206

Table 3: Sizes of the datasets per crisis case.

columns. These labels corresponded to the output of the transformer, the output of the SVM and a randomly assigned label. The judges were not informed about the systems that were used, and the order of the systems was different for each judge.

The judges were asked whether they agreed, disagreed, agreed with doubt or were unable to judge the tweet's emotional value by color coding the labels (green if they agreed with the label, red if they disagreed, orange for a doubtful agree and grey if they were unable to judge the tweet's emotional value). It was possible to agree with different labels for one sentence (e.g., agreeing both with anger and sadness or joy and love). If they disagreed, they had to fill out the corrected label in the fifth column. A translated example of the set-up of this validation experiment is shown in Figure 3.

We then calculate the acceptance rate (i.e., the proportion of items for which a judge agrees with the output), by summing up the items for which they agreed or doubtfully agreed, and dividing this number by the total number of assessable instances (i.e., those they did not indicate as impossible to judge). With A the number of 'agrees', D the number of 'doubtful agrees' and N the number of 'not agrees', this is formalized as follows:

$$\text{Acceptance rate} = \frac{\text{A} + \text{D}}{\text{A} + \text{N} + \text{D}}.$$

Tweet	**System 1**	**System 2**	**System 3**	**Correction**
There's water everywhere. It's extremely high now. The question is. Whether the floodbanks will hold. We can't go home until Sunday at least. So can you please keep your fingers crossed for us? ❤️ #goodmorning #evacuation #hug #floodingLimburg #maas #Roermond #eilandgo	sadness	love	joy	fear
#floodingLimburg @Person did you not notice that a disaster has occurred in #Limburg? Too far from Amsterdam? "I'm here for all Dutch people" remember? (within a radius of 50km?)	anger	fear	neutral	
You anticipated that well! #flooding	love	neutral	joy	

Figure 3: Translated example of the validation experiment set-up.

Application testing procedure

In addition to validating emotion detection models (SQ2), we propose a framework for using such models to monitor the emotional climate on social media. Following Marx et al. (2021), a first step in emotional climate monitoring on social media is to analyze the temporal evolution of emotions during a crisis and identify peak phases. Additionally, we suggest to identify the key topics that are associated with the found emotions, and finally, analyze which influence emotional content has on social media traffic (i.e., how emotional content is associated with the number of likes, retweets, responses, etc.). In the second experimental part of this paper, we will thus test the emotion detection model based on the steps proposed in this framework.

Based on the findings of the validation study, we will proceed with either the transformer or SVM. The selected model will be applied to the crisis-related tweets described above. The model calculates probability scores for all emotion categories (all scores summing to 1), after which the category with the highest probability is outputted as predicted label by the model.[7] The probabilities and label predictions will be used in different steps of the application testing procedure.

Temporal evolution First, the total number of tweets and the counts for each predicted emotion label are plotted per day, this to get a general overview of the tweet counts and emotional fluctuations during the crisis period. To get a better grasp of the emotional fluctuations relative to each other, we also plot the fluctuations in the proportional share of each emotion category. Instead of the predicted labels, we use the label probabilities for this second graph in order to keep as much information as possible. At each day, we calculate the sum of probabilities of a specific emotion label across all instances of that day, and divide that by the number of instances of that day: $P(E)_t = \frac{\sum_{i=1}^{N_t} P(E_i)}{N_t}$, with N_t being the number of instances on day t and $P(E_i)$ being the probability of emotion E for instance i. By looking at the proportional instead of absolute frequencies, classification mistakes have less impact and we can more easily interpret distributional changes.

To objectively measure changes in emotion distribution, we will make use of a peak detection algorithm.[8] We will detect peaks based on the princi-

[7]As transformer models generally tend to produce well-calibrated probabilities, we have not explicitly assessed the calibration of the probabilities, nor have we used any calibration techniques to correct the probabilities.

[8]Implementation from https://stackoverflow.com/questions/22583391/peak-signal-detection-in-realtime-timeseries-data.

ple of statistical dispersion: if a datapoint (in our case: the relative emotion frequency on a specific day) is a specified *n* standard deviations away from the mean, the algorithm signals a peak. Ideally, the mean would be based on a pre-crisis reference period. However, as we do not have enough pre-crisis historical data at our disposal, we opt for a moving mean, where the mean is calculated on smoothed data in a moving time window. The number of days in the time window is then given as a parameter. A final parameter in the implementation of the algorithm is a value related to the influence of detected peaks and troughs on the threshold. If this parameter is set to 0, detected peaks and troughs have no influence on the mean and standard deviation; if it is set to 1, data points leading to peaks and values are treated as normal data points. By way of illustration, we will apply this algorithm on the COVID-19 dataset, using a threshold of 3 standard deviations above a moving mean with a time window of 10 days. We set the last influence parameter to .1, keeping the influence of peaks and values low.

Topics Besides keeping a finger on the pulse of the emotions on social media during crisis times, crisis researchers or professionals might also be interested in the specific targets these emotions are associated with. Aspect-based emotion analysis, analogous to aspect-based sentiment analysis or ABSA (Pontiki et al., 2015), would solve this task. Such systems traditionally consist of three subtasks: aspect extraction (identifying the target of the emotion), aspect classification (labeling the targets with pre-defined aspect classes) and sentiment or emotion classification. Because of the nature of this task, ABSA systems are highly domain-dependent, so versatile open-source tools are currently not available.

One could turn to named entity recognition instead, for which many off-the-shelf tools exist. Although such an approach gives more insights in the important players in a crisis (see Table 14 in the Appendix for an overview of the 5 most frequent entities found by the spaCy[9] entity recognizer for each of the four crises), the limitation to entities is too narrow. This is especially clear in the COVID case, where emotion aspects are often not named entities, but concepts or objects like *face masks*, *lockdown* or *vaccination*.

A more comprehensive technique to identify aspects is topic modeling. Although it does not extract the aspect from the text but classifies the complete document as a topic – from a set of topics found in an unsupervised manner –, this approach is able to pinpoint both entities and other concepts and objects that are being mentioned.

We use BERTopic to this purpose, a state-of-the-art topic modeling tech-

[9]https://spacy.io/api/entityrecognizer

nique based on neural networks (Grootendorst, 2022).[10] BERTopic utilizes pre-trained embeddings from transformers to capture the semantic meaning and context of words within a corpus, and transforms the documents to numerical representations. Then, the dimensionality of these representations is reduced by applying the UMAP dimension reduction technique. Next, the density-based clustering technique HDBSCAN is used to identify clusters (i.e., topics) of documents. For each of the topics, a bag-of-words representation is then made. From these bag-of-words representations, class-based TF-IDF (c-TF-IDF) is called to determine the topic-specific words. The combination of transformer-based embeddings and c-TF-IDF makes that BERTopic shows state-of-the-art results, outperforming other approaches like LDA or STM (Chen et al., 2023; Krishnan, 2023; Meaney et al., 2022).

For each crisis case, we fitted a separate topic model. Following the recommendations from BERTopic, we do not preprocess the texts, apart from removing the hashtags that were used as search terms for collecting the data. We use multilingual BERT (Devlin et al., 2019) as embedding model and the default parameters. Only for the COVID-19 case, we adapt the parameters for the number of clusters (normally set to 'auto', which means that the model determines the topic size itself) and minimum topic size (default is 10), to '100' and '300' respectively, this to encourage the model to make fewer but larger clusters for this very large subset.

We then investigate whether there is a relation between the derived topics and their emotional value, by performing five regression analyses (one for each emotion) per crisis case. The emotion probability found by the transformer model is used as the response variable and the topics (transformed to dummy variables) as predictor variables. This allows us to identify key topics that are associated with specific emotions.

Traffic Finally, we analyze the impact of emotions on the social media traffic. We extract the number of retweets, replies, likes and quotes that every tweet received from the metadata of that tweet and link that to the predicted emotion labels. Then, we perform regression analyses with the emotion probabilities as predictor variables and the number of likes/quotes/replies/retweets as response variable.

[10]https://maartengr.github.io/BERTopic/index.html

Results

Validation Results

The output of the transformer, the SVM and a random baseline on 120 instances of each crisis case was judged by three linguists. An overview of the judgment results is shown in Table 4. What stands out immediately is that there is quite some variability in what the participants labeled as 'impossible to judge'. Whereas the first judge assigned this label 64 times, the second judge only used it for 15 instances. The number of items on which the judges agreed with the output of the systems, however, is more equal across judges.

Judge 3 agrees least with any of the system outputs. However, all judges have a clear preference for the transformer model, compared to the SVM and random classifier. We quantify this by calculating the acceptance rate, which we define as the proportion of agrees (full agrees and doubtful agrees) with respect to the numbers of all assessable items (all items except those impossible to judge). This proportion ranges between 47% and 66% for the transformer, between 28% and 45% for the SVM and between 22% and 33% for the random classifier.

We provide confusion matrices for the transformer and SVM in Figures 4 and 5 respectively. We only take into account the items that were deemed assessable by all three judges and for which at least two judges provided the same emotion category. This majority vote was considered the gold label and resulted in a set of 384 instances. Based on these gold labels, we also calculate precision, recall and F-score per emotion for both the transformer and SVM model (Table 5).

F-score is highest for the transformer model for all emotion categories.

Judgment	**Judge 1**			**Judge 2**			**Judge 3**		
	T	**SVM**	**R**	**T**	**SVM**	**R**	**T**	**SVM**	**R**
Agree (A)	**221**	139	102	**240**	169	100	**208**	121	98
No agree (N)	141	229	279	206	286	356	234	321	344
Doubtful agree (D)	54	48	35	19	10	9	1	1	1
Impossible to judge (I)	64	64	64	15	15	15	37	37	37
Acceptance rate	**66%**	45%	33%	**56%**	38%	23%	**47%**	28%	22%

T = EmotioNL transformer model, SVM = support vector machine, R = random output.
Model with highest number of labels for which the judges agreed with the output is in bold.

Table 4: Results of the validation experiment.

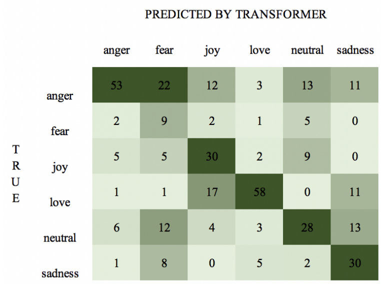


Figure 4: Confusion matrix of the transformer model.

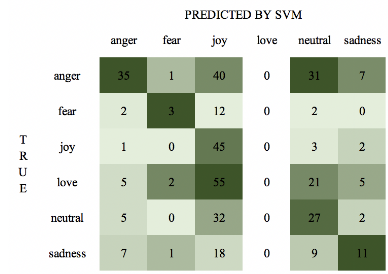


Figure 5: Confusion matrix of the SVM model.

Only the fear precision and joy recall are higher in the SVM. The transformer model seems to miss many anger instances (recall = 47%), even though it is the most common class, and confuses them with fear as illustrated in Figure 4. However, these confusion matrices create a distorted image, because fear is actually only rarely predicted (see infra in the section 'Application testing results'). In the setup of this experiment, however, we started from the same number of predictions for every emotion category, which results in an overrepresentation of certain categories. On the dataset as a whole, we thus assume that the anger-fear confusion is less problematic. On the other hand, if the system predicts anger, there is a high chance that this label is indeed correct (precision = 77%).

The presence of the neutral class is underestimated as well, and is often confused with negative emotions. Overall this emotion has a rather low precision (49%) and recall (42%). Furthermore, sadness is often confused with anger, neutral and remarkably also with love (probably in the sense of pity).

This validation study thus shows that the transformer model yields the most satisfying results and corroborates the intrinsic evaluation in which the transformer model also came out as best performing model. Though there is quite some variability among the participants which labelled the validation set, the acceptance rate of the transformer model can go up to 66%, while it does not go above 45% for the SVM.

We believe that the lower computational cost of the SVM does not make up for such a notable difference in performance (both in this extrinsic validation experiment, but also in the intrinsic evaluation). Moreover, the availability of open-source fine-tuned models makes it possible to employ emotion detection models in an almost off-the-shelf manner, without the explicit

Emotion	Transformer			SVM		
	Precision	Recall	F	Precision	Recall	F
Anger	0.779	0.465	0.582	0.636	0.307	0.414
Fear	0.158	0.474	0.237	0.429	0.158	0.231
Joy	0.462	0.588	0.517	0.223	0.882	0.356
Love	0.806	0.659	0.725	0	0	0
Neutral	0.491	0.424	0.455	0.290	0.409	0.340
Sadness	0.556	0.652	0.600	0.407	0.239	0.301

Table 5: Precision and recall per emotion category for the transformer model.

need of GPUs. Apart from EmotioNL, which is a Dutch model, fine-tuned transformer-based emotion detection models exist for other languages as well, including for English (Barbieri et al., 2020)[11].

We do advise to first validate a model (like in this experiment) before using it in applied settings. Researchers can then decide whether the model performance is acceptable enough to use in practice and pinpoint challenges which can be taken into account in the analysis. Overall, we should critically interpret frequencies of predicted emotions. We should thus avoid making claims where we, for example, compare the number of angry and sad tweets. Instead, it is more informative to identify significant fluctuations in emotion distributions.

Application testing results

In this section, we test the application of the EmotioNL transformer model. We therefore rely on our proposed emotional climate monitoring framework for social media, consisting of three steps dealing with a) Temporal evolution, b) Topics, and c) Traffic.

A. Temporal evolution

In order to get a general overview of the emotional fluctuations during a crisis period, we use the transformer model to predict the emotional value of each instance in the crisis datasets and plot the daily count per emotion category alongside the total number of tweets per day.

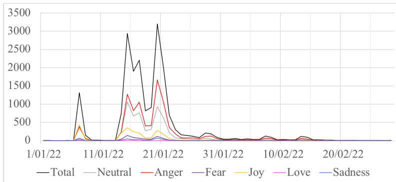
[11] https://huggingface.co/cardiffnlp/twitter-roberta-base-emotion

As seen in Figure 6, the emotion lines follow the total tweet count in each crisis case rather strictly. The biggest peaks in these lines are directly correlated to the emergence of the crises: in the graph belonging to The Voice of Holland (Figure 6a), the first peak on 7/1/2022 coincides with the broadcasting of the first episode of the 12th season, while the peaks of 15/1/2022 and 20/1/22 refer to the announcement that the season would be suspended indefinitely due to the allegations against crew members of the show, and the broadcasting of the BOOS episode in which the allegations were raised, respectively.

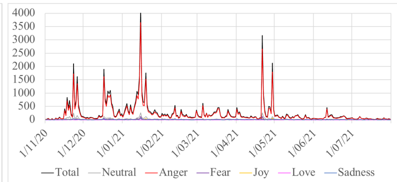
For the childcare benefits scandal (Figure 6b), many peaks can be observed, of which the biggest ones are those on 23/11/2020 (in the middle of the investigation by the Parliamentary Interrogation Committee, in which it appears that the financial suffering of parents due to the strict policy had been known for many years and in which it was decided that victims should be allocated victim support); on 17/12/2020 (the report of the Parliamentary Interrogation Committee is published, which concludes that the principles of the rule of law have been violated and that both the ministries and the judiciary have contributed to this); on 15/1/2021 (the cabinet resigns because of the childcare benefits scandal); on 21/4/2021 (in the Council of Ministers it comes to light that the cabinet deliberately withheld information about the childcare benefits scandal) and, finally, on 29/4/2021 (The House of Representatives is debating the fact that the government has deliberately withheld information about the childcare benefits scandal. By then the payment of compensations is failing, while it was promised that it would be sorted before May 1st).

In Figure 6c, the peak on 15-16/7/2021 is a direct consequence of the floods between 13/7/2021 and 15/7/2021. Regarding COVID-19 (Figure 6d), a first peak is visible on 27/2/2020, when the first COVID infection was reported in the Netherlands, and a second, much larger peak on 12/3/2020, the day that the lockdown (which would start 13/3) was announced.

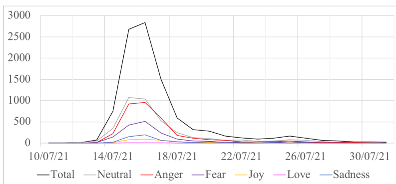
As we know the model still makes wrong predictions (see previous section), it is important to not focus too much on absolute emotion frequencies, but rather gauge fluctuations in the proportional share of each emotion category. The plots in Figure 7 reveal the proportional frequency lines of each emotion category. Note that at each timestamp, the values of all emotion lines sum to 1. Peaks and troughs that were not visible in the subplots in Figure 6 because of low absolute frequencies, do pop out here. In the joy line for The Voice of Holland (7a), for example, there is a trough on 12/1/2022 (that was a peak when looking at absolute frequencies) and a peak on 5/2/2022.



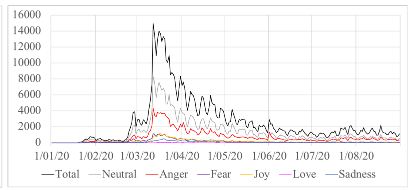
(a) The Voice of Holland.



(b) Childcare benefits scandal.



(c) Floods.

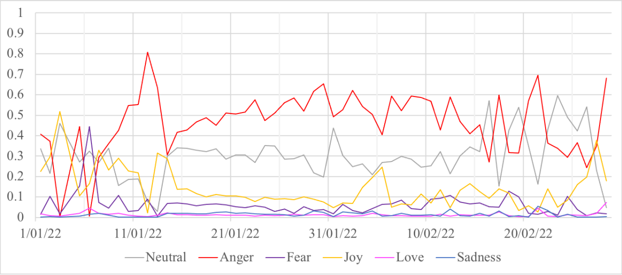


(d) COVID-19.

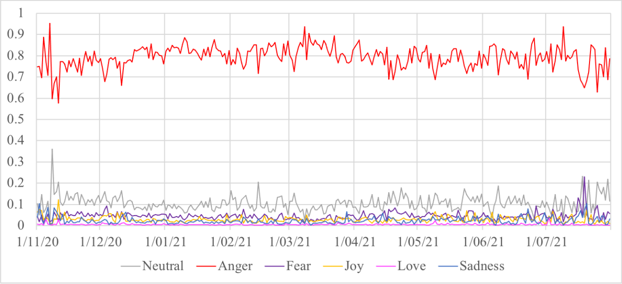
Figure 6: Absolute tweet frequencies in total and per emotion category for each crisis case.

Some general observations can be made when comparing the emotional distributions of the different crises. The proportion of anger in the tweets about the child benefits scandal clearly stands out. The emotions in the tweets about the floods are a bit more evenly distributed, while the ones about COVID and The Voice of Holland are primarily dominated by neutral and angry tweets. One way to explain this could be the crisis responsibility that is attributed to certain stakeholders in these crisis situations. The childcare benefits scandal and The Voice of Holland are intentional crises in which there are clear actors to blame. In a natural disaster like the floods of 2021, that is much less the case. Although the COVID pandemic is not an intentional crisis either, there are some actors that do have a large responsibility with which people can disagree, like the government. In SCCT, three clusters are distinguished based on crisis responsibility, namely the victim cluster (an organization is the victim of a crisis), an accident (actions of an organization unintentionally lead to a crisis) and an intentional cluster (an organization knowingly caused a crisis). Note, however, that SCCT is mainly used in an organizational context and thus less applicable to natural disasters or health crises. We thus do not use this typology of crisis types, but merely wish to mention the notion of attribution in this context as a possible explanation.

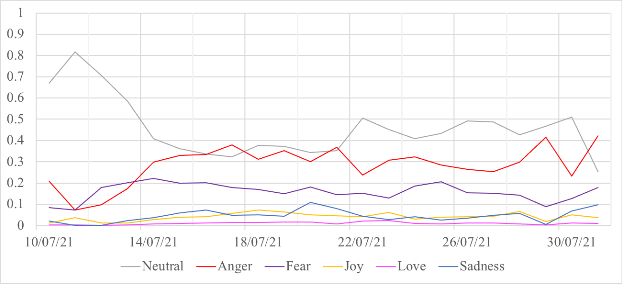
The main focus, however, should be on identifying significant fluctuations in emotion frequencies over time. To objectively measure these fluctuations, but also to use emotion detection as a monitoring tool in real-time data streams, a peak detection algorithm can be used.



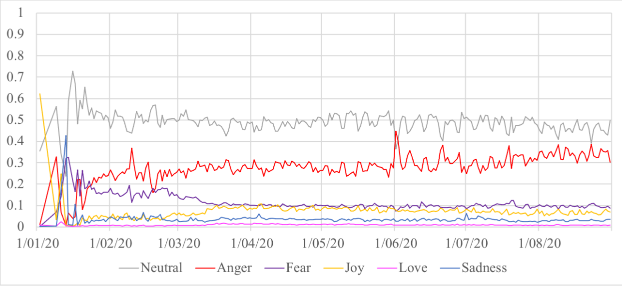
(a) The Voice of Holland.



(b) Childcare benefits scandal.



(c) Floods.



(d) COVID-19.

Figure 7: Relative tweet frequencies per emotion category for each crisis case.

When the algorithm is applied on the COVID-19 dataset, several flags are signalled. These are shown in Figure 8. Ideally, interpret the flags together with the frequency graphs to have an idea of the order of magnitude. The first flags appear in the beginning of February, namely for fear, anger and sadness. In this period, the number of infections and deaths increased in China, and the first rumours about regulations in the Netherlands went around (see Examples 1 and 2 in Table 6). However, when inspecting the tweets that were labeled with 'fear' by the emotion detection system, another event popped up, namely storm Ciara (Example 3). Also in mid March some peaks are detected, which can be related to the first lockdown and regulations. Remarkable is that especially the positive emotions peak in this period. Probably, people were still optimistic about the pandemic and lockdowns at this stage, and many tweets were posted to encourage each other (Example 4 and 5). Also in the beginning of April, a peak in positive emotions can be observed. This is probably linked to the start of the spring vacation in Belgium and the nice weather that weekend, as illustrated in Example 6.

These analyses show that the emotion fluctuations and peaks based on the predictions by the EmotioNL model, are linked to events that occurred during the crisis. We therefore believe that emotion detection models can, despite their imperfect performance and when used with knowledge about

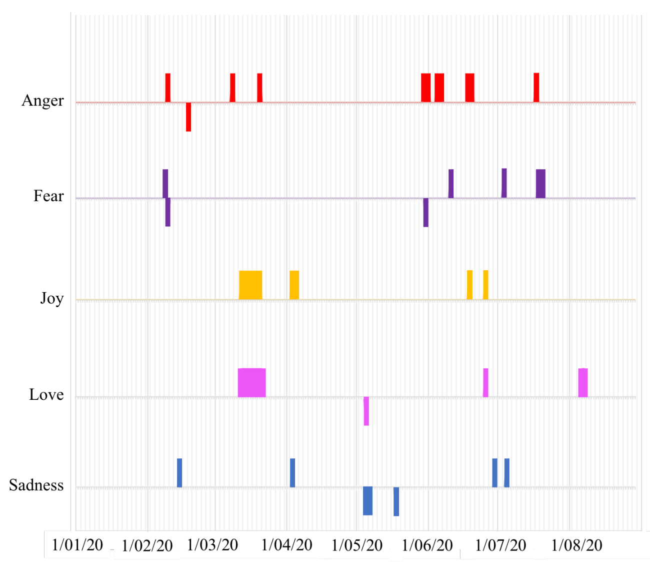


Figure 8: Illustration of flags in monitoring tool for the COVID-19 case.

	Tweet
(1)	15/02/2020 – 1526 people died from the #CoronaVirus – Sadness
(2)	09/02/2020 – Thread #CoronaVirus "The indications that this is a very serious epidemic are mounting. The official numbers are rubbish." "We're talking about the largest quarantine in history + restricted freedom of movement" – Fear
(3)	09/02/2020 That you think being locked up on a ship with the coronavirus is the worst thing that could happen to you #StormCiara – Fear
(4)	15/03/2020 every disadvantage has it's advantage. Climate goals will be completely fine. Thanks to #coronavirus . CO2 emissions in China fell by no less than 25%. The rest of the world now follows. Greta can clap her hands. – Joy
(5)	15/03/2020 Look after each other a little. That's how we get there. And some extra support for all nurses and health care providers who continue because it is their profession and calling to take care of weaker children and the elderly. #Corona – Love
(6)	04/04/2020 – 'Home sweet home.'. A lovely sunny weekend, we stay home. Enjoying the garden, cleaning up and cuddling with the girls. What are your plans? #family #stayhome #corona – Joy

Table 6: Translated examples from the COVID-19 dataset. Original tweets are shown in Table 13 in the Appendix.

their flaws, can be informative for crisis communication practitioners and researchers interested in the temporal evolution of the emotional climate on social media.

B. Topics

Using BERTopic, we fitted a topic model for each crisis case. We identified 80 topics for The Voice of Holland, 77 for the childcare benefits scandal, 40 for floods and 148 for COVID-19. A list of topics per crisis case is shown in Tables 19-22 in the Appendix.

We notice that various important players are identified by the models. In The Voice of Holland for example, we see topic clusters concerned with the accused coaches (Topic 1 and 5) and band leader *Jeroen Rietbergen* (Topic 26 and 68), but also for *Tim Hofman* who raised the issue in his BOOS documentary (Topic 40) and others involved, such as coach *Anouk* (Topic 8) and *Linda De Mol* (Topic 3). The COVID-19 topic model exhibits clusters concerning the *health care* sector (Topic 3), *face masks* (Topic 15), *vaccination* (Topic 16), *teleworking* (Topic 18) and *lockdown* (Topic 30).

In the floods and childcare benefits models, the distribution of topics is less diverse. In the former, most tweets are classified under the topics concerning the location of the disaster (Topic 0, 2 and 3). However, we do observe some interesting other topics like *climate change* (Topic 1), *media* (Topics 7) or *government* (Topic 13). In the childcare benefits model, the most prominent topic is clearly the one concerning the *Rutte cabinet* (Topic 0). Other interesting topics are, for example, the one concerning the *victims* (Topic 2), the *Tax and Customs Administration* (Topic 3), *media* (Topic 7) and the political party *CDA* and members thereof (Topic 8).

We investigate whether there is a relation between the derived topics and their emotional value, by performing five regression analyses (one for each emotion) per crisis case. We find that, even with such a high number of predictor variables (i.e., a dummy variable for each identified topic), all models are significant (although accounting only for a minimal proportion of the variance). In fact, many topics significantly contribute to the emotional value in tweets. Table 7 lists which five topics contribute most (only in the positive direction) to each emotion category for each of the four crisis cases.

For The Voice of Holland, the topics contributing most to anger and sadness overlap quite a bit and relate to *abuse of power* and *victims*. Also the top 5 for joy and love overlap and mostly concern specific candidates of the program. There is even one topic that specifically emerged about the word 'prachtig' (wonderful). For fear, the most contributing topic is related to the winner of the contest, which suggests that the connotation of fear is closer to excitement in this case. There are also some topics with high contribution that cannot be immediately linked to The Voice of Holland, e.g. *racism* for anger or *Russia-Ukrain* and *coronavariant* for fear.

For the childcare benefits scandal, we mainly see that the topics in anger and sadness are related to the scandal (e.g., topics like *Rutte*, *xenophobia*, *liar*, *victims*, *recovery operation*). However, for the other emotion categories, the link to the childcare benefits scandal is not so straightforward. This might be explained by the clear dominance of the anger category in this crisis case.

For floods, we again see overlap in the topics that contribute to joy and love, and mostly, these topics are not related to the crisis (e.g., *songs* and *pictures*), although *armed forces* is an important topic here, and refers to their help in the disaster. For fear, we find only four positively contributing topics in total, of which three more generally concern the floods (the ones in Belgium and the Netherlands, but also the ones in China that emerged around the same time), and one unrelated topic (the death of Dutch crime reporter Peter R. De Vries). The latter was also the topic with the highest contribution to sadness. Other topics contributing to this crisis were again

	The Voice of Holland	Childcare Benefits Scandal
Anger	Topic 49 abuse of power	Topic 38 xenophobia
	Topic 57 *	Topic 61 Rutte
	Topic 63 racism	Topic 54 liar
	Topic 42 victims	Topic 42 fascists
	Topic 68 Jeroen Rietbergen	Topic 27 Christmas
Fear	Topic 74 winner	Topic 73 good news
	Topic 47 Russia Ukraine	Topic 39 2020-2021
	Topic 23 Maan, Marco Borsato	Topic 59 Elfstedentocht
	Topic 10 *	Topic 58 MH17
	Topic 58 corona variant	Topic 52 Alzheimer
Joy	Topic 52 Jefferson	Topic 73 good news
	Topic 43 Simon	Topic 44 curfew
	Topic 12 wonderful	Topic 74 Christmas
	Topic 38 Spotify	Topic 63 movie clip
	Topic 79 Glennis	Topic 67 bicycles
Love	Topic 12 wonderful	Topic 74 Christmas
	Topic 64 Glennis	Topic 45 restore trust
	Topic 67 football	Topic 39 2020-2021
	Topic 79 Glennis	Topic 26 Christmas
	Topic 8 Anouk	Topic 73 good news
Sadness	Topic 9 victims perpetrators	Topic 74 Christmas
	Topic 69 family De Mol	Topic 64 pension crisis
	Topic 20 *	Topic 70 recovery operation
	Topic 49 abuse of power	Topic 52 Alzheimer
	Topic 10 *	Topic 2 victims
	Floods	**COVID-19**
Anger	Topic 22 rubberneckers	Topic 61 racism
	Topic 39 PCR tests	Topic 96 ceasefire
	Topic 13 government	Topic 7 politics
	Topic 38 racism	Topic 77 Halsema
	Topic 5 trending on Twitter	Topic 82 protesters
Fear	Topic 20 Peter R. de Vries	Topic 93 angst
	Topic 8 disaster	Topic 20 corona crisis
	Topic 31 Flooding China	Topic 115 quarantine
	Topic 0 Limburg	Topic 56 economy
		Topic 52 pandemic
Joy	Topic 32 song	Topic 92 humor
	Topic 17 donation	Topic 76 flowers
	Topic 11 king	Topic 111 birthday
	Topic 25 armed forces	Topic 139 garden
	Topic 18 picture	Topic 21 music
Love	Topic 32 song	Topic 76 flowers
	Topic 18 picture	Topic 96 ceasefire
	Topic 25 armed forces	Topic 111 birthday
		Topic 83 volunteers
		Topic 132 wedding
Sadness	Topic 26 Peter R. de Vries	Topic 127 funeral
	Topic 36 donation	Topic 116 cancer
	Topic 3 Belgium Germany	Topic 142 homeless
	Topic 0 Limburg	Topic 28 elderly
		Topic 109 disabled

* Nonmeaningful topics are left out.

Table 7: Top 5 most (positively) contributing topics per emotion category for each crisis case.

more general and related to the location of the disaster. For anger, however, we do see some more specific topics, like *rubberneckers* (people who look for sensation in disaster situations without helping) and *government*. Again, we find some unrelated topics like *PRC test* and *racism*.

In the COVID-19 data, the fear topics are clearly related to the disease, with topics as *corona crisis*, *quarantine* and *pandemic* in the top 5. Interestingly, one topic is even completely centered around the word 'angst' (fear). Another remarkable fear topic is *economy*, probably concerning people's fear for the negative impact of the corona crisis on the economy. For sadness, we see the topics *cancer*, *homeless*, *elderly* and *disabled* as top 5 topics, indicating the most vulnerable groups in this crisis. *Funeral* is the most contributing topic, and could either relate to the deaths of the pandemic, or to the fact that funerals could not be organized in a normal way due to the corona restrictions. For joy and love we see many general positive topics, like *humor* (probably a coping strategy for many during the crisis), *flowers*, *birthday*, *wedding*, *garden* and *music*. Two important topics are *volunteers* and *ceasefire*, the latter one referring to the global ceasefire proposed by United Nations Secretary-General António Guterres on 23 March 2020, as part of the United Nations' response to the COVID-19 pandemic. Remarkably, this topic was also in the top 5 topics contributing to anger, next to topics like *politics*, *Halsema* (the mayor of Amsterdam), *protesters* and the more general topic *racism*.

In general, we can conclude that specific topics significantly contribute to specific emotion categories identified in the tweet. Topics that contribute more to anger are mostly related to entities which could be attributed a certain degree of blame (e.g., *rubberneckers* in floods or *Rutte* in the childcare benefits scandal), while topics contributing to sadness are more related to victims (e.g., *elderly* in COVID-19). Topics that contribute most to joy and love are often not directly related to the crisis (e.g., *flower* in COVID-19), or had a positive role in the crisis (e.g., *armed forces* in floods or *volunteers* in COVID-19). Based on these insights, we believe that emotion detection models, combined with topic modeling, can be a valuable tool in the topic-related step of emotional climate monitoring.

Link between emotions and traffic

Finally, we look at the relation between emotions and social media traffic. To investigate this, we extract the number of retweets, replies, likes and quotes that every tweet received from the metadata of that tweet.

Table 8 shows the average number of likes per tweet, grouped by emotion

Emotion	The Voice of Holland	Childcare Benefits	Floods	COVID-19
neutral	6.93	10.08	10.56	3.73
anger	8.73	10.22	11.54	**7.12**
fear	7.87	9.42	8.43	3.84
joy	7.55	10.41	**17.14**	6.80
love	**21.04**	**14.56**	**17.97**	**8.17**
sadness	9.69	**16.21**	12.85	**7.74**

Table 8: Average number of likes per tweet, grouped by emotion category.

category. An overview of the other traffic indicators is given in Tables 15–18 in the Appendix.

It immediately stands out that tweets classified as love receive most likes. This is especially clear in The Voice of Holland, but also in the floods case tweets classified as love receive most likes (together with joy). In the COVID-19 case there is a smaller difference between the categories. Again, tweets classified as love receive most likes, followed by angry and sad tweets. In the childcare benefits affair, sadness and love receive most likes.

The higher number of likes for tweets classified as love can be explained by the exceptionality of expressing positive emotions during negative events. This is in line with findings that news is more likely shared when it is positively framed compared to when it is negatively framed (van der Meer & Brosius, 2022), although counterexamples were found in many other studies as well (Rathje et al., 2021; Robertson et al., 2023).

We statistically verify these tendencies by performing regression analyses. As the number of likes highly exceeds the other traffic indicators, we only report the results of the analysis of likes (results in Tables 9-12), but similar tendencies were found for the other traffic indicators. Although the models were significant for all crisis cases except for The Voice of Holland, the extremely low R^2 values indicate that emotion explains only a very small part of the variation in these models. However, the models do indicate similar tendencies as discussed above, namely that love and anger lead to the most traffic in The Voice of Holland, love and sadness in the childcare benefits scandal, and joy in the floods case. For the COVID-19 case, all emotion categories significantly contribute to the received likes, with fearful tweets leading to fewer likes and angry, loving, joyful and sad tweets to more likes.

Seeing the low R^2 values, emotion seems to have only a minimal impact on social media traffic, at least in the four crisis cases studied in this paper.

	R^2 = 0.0002976, F(5,18496) = 2.101, p = 0.06212				
	B	**Std. Error**	**T**	**P**	
(Intercept)	6.518	0.973	6.696	2.2e-11	***
Anger	2.548	1.262	2.018	0.0436	*
Fear	-1.960	3.414	-0.574	0.5660	
Joy	-0.492	2.165	-0.227	0.8202	
Love	33.830	14.925	2.267	0.0234	*
Sadness	3.001	4.650	0.645	0.5187	

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Table 9: Regression analysis results The Voice of Holland.

	R^2 = 0.0001975, F(5,66955) = 3.645, p = 0.002677				
	B	**Std. Error**	**T**	**P**	
(Intercept)	9.970	1.365	7.303	2.84e-13	***
Anger	0.274	1.439	0.190	0.84910	
Fear	-3.805	3.315	-1.148	0.25113	
Joy	-5.875	4.158	-1.413	0.15773	
Love	57.447	20.997	2.736	0.00622	**
Sadness	9.310	3.825	2.434	0.01493	*

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Table 10: Regression analysis results Childcare Benefits.

	R^2 = 0.0008681, F(5,9917) = 2.724, p = 0.01826				
	B	**Std. Error**	**T**	**P**	
(Intercept)	9.569	1.274	7.510	6.42e-14	***
Anger	2.700	1.877	1.439	0.1502	
Fear	-1.789	2.919	-0.613	0.5400	
Joy	12.760	5.468	2.333	0.0196	*
Love	13.658	16.883	0.809	0.4185	
Sadness	0.986	3.785	0.260	0.7946	

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Table 11: Regression analysis results Floods.

	R^2 = 0.001467, F(5,609200) = 180, p < 2.2e-16				
	B	**Std. Error**	**T**	**P**	
(Intercept)	3.205	0.112	28.516	< 2e-16	***
Anger	4.868	0.191	25.430	< 2e-16	***
Fear	-1.170	0.396	-2.953	0.00314	**
Joy	2.335	0.392	5.951	2.67e-09	***
Love	19.831	2.009	9.872	< 2e-16	***
Sadness	5.385	0.559	9.636	< 2e-16	***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Table 12: Regression analysis results COVID-19.

Therefore, the traffic step in the emotional climate monitoring framework might be less essential for crisis managers. However, seeing the diversity in emotion-traffic relations across crisis cases (and the conflicting findings between researchers in previous studies about engagement and emotional framing), we do believe emotion detection models can be a helpful tool for crisis communication researchers to delve deeper into this question.

Conclusion

In this paper, we investigated how emotion detection models can aid research on crisis communication and social media. We focused mainly on the Dutch transformer-based model EmotioNL, but also included a less computationally costly SVM model based on n-grams and dictionary features. The models were not specifically developed for the crisis domain, as they were trained on the EmotioNL background corpus for identifying the fine-grained emotion categories anger, fear, joy, love, sadness and neutral. We were mainly interested in how these models would perform in the context of crisis communication when employed in an out-of-the-box manner (without further fine-tuning).

To gather relevant data for evaluating the models in the crisis domain, we collected Dutch tweets concerning four recent and diverse crisis cases in the Netherlands, namely the scandal in the singing competition The Voice of Holland, the childcare benefits scandal, the floods of Summer 2021, and the COVID-19 crisis.

First, we performed a validation experiment, in which we applied the transformer-based model and the SVM to a subset of the crisis tweets and compared the output of the models. Three judges were presented with the output of both machine learning systems and labels from a random classifier. Although there was a high variability in the assessments of the judges, the transformer model was considered best by all judges, reflected in an acceptance rate between 47% and 66% (compared to only 28 to 45% for the SVM). We therefore suggest to give preference to transformer-based models, but do advise to validate the model before use in applied settings to get more insights into the flaws of the model.

Secondly, we proposed a framework for emotional climate monitoring on social media and assessed the application potential of the transformer model based on the steps in this framework. In the first step of the framework, namely analyzing the temporal evolution of emotions during a crisis, we found that based on the emotion detection model's predictions, peak phases can be identified that correspond to important events during the crisis. This indicates that, even though the performance of the model is not perfect, fluctuations over time in the emotional climate can be captured.

The second step of the framework, i.e., identifying key topics and investigating how these are associated with the emotions found by the model, revealed relevant topic-emotion associations that hinted towards a promising research potential for emotion detection combined with topic modeling.

The last step consisted of assessing the impact of emotions on social media traffic. Although emotions seem to have minimal impact on social media traffic overall, there was a significant relation between the emotion found in tweets and the number of likes those tweets received, and these relations differed across crisis cases. Therefore, we do believe that emotion detection models can be used by crisis communication researchers to delve deeper into the question how specific emotions conveyed in messages impact social media engagement.

In conclusion, this research highlights the potential of emotion detection models, particularly transformer-based ones, in enhancing our understanding of crisis communication on social media. Crisis communication researchers and practitioners can turn to the emotional climate monitoring framework proposed in this paper, but even when relying on established frameworks like SCCT, ICM or SMCC, we believe there is a high potential for transformer-based emotion detection models. A recurrent idea in these frameworks is that crisis responses should be tailored to the public's emotions, so automating the process of analyzing these emotions would be highly beneficial. Moreover, we showed that emotion detection models that are not specifically developed for crisis communication contexts can still offer valuable insights. This opens up possibilities, as many emotion detection models have been made publicly available, even for other languages than Dutch or English.

Acknowledgements

This research was funded by Research Foundation–Flanders under a Strategic Basic Research fellowship with grant number 3S004019.[12] The research was part of a research stay at ASCoR, for which a travel grant was awarded by Research Foundation–Flanders.

References

Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for text-based emotion detection: A review of BERT-based approaches. *Artificial Intelligence Review*, *54*(8), 5789–5829.

Austin, L., Fisher Liu, B., & Jin, Y. (2012). How audiences seek out crisis information: Exploring the Social-Mediated Crisis Communication model. *Journal of applied communication research*, *40*(2), 188–207.

Barbieri, F., Camacho-Collados, J., Espinosa Anke, L., & Neves, L. (2020). TweetEval: Unified benchmark and comparative evaluation for tweet classification. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 1644–1650. https://doi.org/10.18653/v1/2020.findings-emnlp.148

Boot, P., Zijlstra, H., & Geenen, R. (2017). The Dutch translation of the Linguistic Inquiry and Word Count (LIWC) 2007 dictionary. *Dutch Journal of Applied Linguistics*, *6*(1), 65–76.

Boukes, M., Van de Velde, B., Araujo, T., & Vliegenthart, R. (2020). What's the tone? Easy doesn't do it: Analyzing performance and agreement between off-the-shelf sentiment analysis tools. *Communication Methods and Measures*, *14*(2), 83–104.

Buechel, S., Rücker, S., & Hahn, U. (2020). Learning and evaluating emotion lexicons for 91 languages. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 1202–1217. https://doi.org/10.18653/v1/2020.acl-main.112

Chaffar, S., & Inkpen, D. (2011). Using a heterogeneous dataset for emotion analysis in text. *Canadian conference on artificial intelligence*, 62–67.

Chaumartin, F.-R. (2007). UPAR7: A knowledge-based system for headline sentiment tagging. *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, 422–425. https://aclanthology.org/S07-1094

Chen, W., Rabhi, F., Liao, W., & Al-Qudah, I. (2023). Leveraging state-of-the-art topic modeling for news impact analysis on financial markets: A comparative study. *Electronics*, *12*(12). https://www.mdpi.com/2079-9292/12/12/2605

[12] https://www.researchportal.be/nl/project/transfer-learning-voor-automatische-emotiedetectie-nederlandstalige-teksten

Choi, Y., & Lin, Y.-H. (2009). Consumer responses to mattel product recalls posted on online bulletin boards: Exploring two types of emotion. *Journal of public relations research*, *21*(2), 198–207.

Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of Situational Crisis Communication Theory. *Corporate reputation review*, *10*(3), 163–176.

Coombs, W. T., & Holladay, S. J. (2022). *The handbook of crisis communication, second edition*. John Wiley & Sons.

De Bruyne, L., De Clercq, O., & Hoste, V. (2021a). Annotating affective dimensions in user-generated content. *Language Resources and Evaluation*, *55*(4), 1017–1045.

De Bruyne, L., De Clercq, O., & Hoste, V. (2021b). Prospects for Dutch emotion detection: Insights from the new emotionl dataset. *Computational Linguistics in the Netherlands Journal*, *11*, 231–255.

De Geyndt, E., De Clercq, O., Van Hee, C., Lefever, E., Singh, P., Parent, O., & Hoste, V. (2022). SentEMO: A multilingual adaptive platform for aspect-based sentiment and emotion analysis. *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, 51–61. https://doi.org/10.18653/v1/2022.wassa-1.5

De Smedt, T., & Daelemans, W. (2012). "Vreselijk mooi!"(terribly beautiful): A subjectivity lexicon for Dutch adjectives. *LREC*, 3568–3572.

Delobelle, P., Winters, T., & Berendt, B. (2020). RobBERT: A Dutch RoBERTa-based Language Model. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 3255–3265. https://doi.org/10.18653/v1/2020.findings-emnlp.292

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186.

Frasnelli, V., Bocchi, L., & Palmero Aprosio, A. (2021). Erase and rewind: Manual correction of NLP output through a web interface. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, 107–113. https://doi.org/10.18653/v1/2021.acl-demo.13

Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Jabreel, M., & Moreno, A. (2019). A deep learning-based approach for multi-label emotion classification in tweets. *Applied Sciences*, *9*(6), 1123.

Jijkoun, V., & Hofmann, K. (2009). Generating a non-English subjectivity lexicon: Relations that matter. *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, 398–405.

Jin, Y. (2010). Making sense sensibly in crisis communication: How publics' crisis appraisals influence their negative emotions, coping strategy preferences, and crisis response acceptance. *Communication Research*, *37*(4), 522–552.

Jin, Y., Liu, B. F., Anagondahalli, D., & Austin, L. (2014). Scale development for measuring publics' emotions in organizational crises. *Public Relations Review*, *40*(3), 509–518. https://doi.org/https://doi.org/10.1016/j.pubrev.2014.04.007

Jin, Y., Liu, B. F., & Austin, L. L. (2014). Examining the role of social media in effective crisis management: The effects of crisis origin, information form, and source on publics' crisis responses. *Communication research*, *41*(1), 74–94.

Jin, Y., Pang, A., & Cameron, G. T. (2010). The role of emotions in crisis responses: Inaugural test of the Integrated Crisis Mapping (ICM) model. *Corporate Communications: An International Journal*, *15*(4), 428–452.

Jonkman, J. G., Boukes, M., Vliegenthart, R., & Verhoeven, P. (2020). Buffering negative news: Individual-level effects of company visibility, tone, and pre-existing attitudes on corporate reputation. *Mass Communication and Society*, *23*(2), 272–296.

Kim, H. S. (2015). Attracting views and going viral: How message features and news-sharing channels affect health news diffusion. *Journal of Communication*, *65*(3), 512–534.

Krishnan, A. (2023). Exploring the power of topic modeling techniques in analyzing customer reviews: A comparative analysis. *arXiv preprint arXiv:2308.11520*.

Liu, B. (2020). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press.

Liu, B.-F., Jin, Y., Austin, L., & Janoske, M. (2012). The Social-Mediated Crisis Communication model: Guidelines for effective crisis management in a changing media landscape. *New media and public relations*, *2*, 257–266.

Lu, Y., & Huang, Y.-H. C. (2018). Getting emotional: An emotion-cognition dual-factor model of crisis communication. *Public Relations Review*, *44*(1), 98–107.

Marx, J., Mirbabaie, M., Czonstke, C., & Stieglitz, S. (2021). Social media analytics and corporate crises-a case study of Boeing's 737 max crashes. *ECIS*.

Meaney, C., Escobar, M., Stukel, T. A., Austin, P. C., & Jaakkimainen, L. (2022). Comparison of methods for estimating temporal topic models from primary care clinical text data: Retrospective closed cohort study. *JMIR Med Inform*, *10*(12), e40102.

Mohammad, S. (2012). Portable features for classifying emotional text. *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 587–591.

Mohammad, S. M. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*.

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, *29*(3), 436–465.

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., Van Schie, K., Van Harmelen, A.-L., De Schryver, M., De Winne, J., & Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior research methods*, *45*(1), 169–177.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic Inquiry and Word Count: LIWC*. Pennebaker Conglomerates.

Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., & Androutsopoulos, I. (2015). Semeval-2015 task 12: Aspect based sentiment analysis. *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, 486–495.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners.

Rathje, S., Van Bavel, J. J., & Van Der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, *118*(26), e2024292118.

Rhodes, J. H., & Vayo, A. B. (2019). The historical presidency: Fear and loathing in presidential candidate rhetoric, 1952–2016. *Presidential Studies Quarterly*, *49*(4), 909–931.

Risius, M., & Akolk, F. (2015). Differentiated sentiment analysis of corporate social media accounts. *Twenty-first Americas Conference on Information Systems*.

Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., & Feuerriegel, S. (2023). Negativity drives online news consumption. *Nature Human Behaviour*, 1–11.

Schoofs, L., & Claeys, A.-S. (2021). Communicating sadness: The impact of emotional crisis communication on the organizational post-crisis reputation. *Journal of Business Research*, *130*, 271–282.

Schuff, H., Barnes, J., Mohme, J., Padó, S., & Klinger, R. (2017). Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus. *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 13–23. https://doi.org/10.18653/v1/W17-5203

Schultz, F., Kleinnijenhuis, J., Oegema, D., Utz, S., & Van Atteveldt, W. (2012). Strategic framing in the bp crisis: A semantic network analysis of associative frames. *Public Relations Review*, *38*(1), 97–107.

Soroka, S. N. (2006). Good news and bad news: Asymmetric responses to economic information. *The journal of Politics*, *68*(2), 372–385.

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). The General Inquirer: A computer approach to content analysis.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American society for information science and technology*, *61*(12), 2544–2558.

Troiano, E., Padó, S., & Klinger, R. (2021). Emotion ratings: How intensity, annotation confidence and agreements are entangled. *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 40–49. https://aclanthology.org/2021.wassa-1.5

Van Atteveldt, W., van der Velden, M. A., & Boukes, M. (2021). The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Communication Methods and Measures*, *15*(2), 121–140.

van der Meer, T. G. (2016). Automated content analysis and crisis communication research. *Public Relations Review*, *42*(5), 952–961. https://doi.org/https://doi.org/10.1016/j.pubrev.2016.09.001

van der Meer, T. G., & Brosius, A. (2022). Credibility and shareworthiness of negative news. *Journalism*.

van Atteveldt, W., & Peng, T.-Q. (2018). When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures*, *12*(2-3), 81–92.

Vignal Lambret, C., & Barki, E. (2018). Social media crisis management: Aligning corporate response strategies with stakeholders' emotions online. *Journal of Contingencies and Crisis Management*, *26*(2), 295–305.

Vilares, D., Peng, H., Satapathy, R., & Cambria, E. (2018). Babelsenticnet: A commonsense reasoning framework for multilingual sentiment analysis. *2018 IEEE symposium series on computational intelligence (SSCI)*, 1292–1298.

Weiner, B. (1986). *An attributional theory of motivation and emotion*. Springer.

Zhu, L., Pergola, G., Gui, L., Zhou, D., & He, Y. (2021). Topic-driven and knowledge-aware transformer for dialogue emotion detection. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1571–1582. https://doi.org/10.18653/v1/2021.acl-long.125

Appendix

Table 13 shows the original tweets of the examples in Table 6. In Table 14, the top five most frequent named entities for each crisis case are given. Tables 15–18 show the average number of traffic indicators (retweets, replies, likes and quotes) per tweet in the four crisis cases, grouped by emotion category. Tables 19–22 show the list of topics found by BERTopic per crisis case.

	Tweet
(1)	1526 mensen overleden aan het #CoronaVirus https://t.co/e6mATz8EVn
(2)	Draadje #CoronaVirus "De aanwijzingen dat dit een heel ernstige epidemie is stapelen zich op. De officiële cijfers zijn rotzooi." "We praten over de grootste quarantaine uit de geschiedenis + ingeperkte bewegingsvrijheid" https://t.co/GztMk4w21R
(3)	Dat je denkt dat opgesloten zitten op een schip met het #coronavirus het ergste is wat je kan overkomen #StormCiara https://t.co/4pC8Fylmhx
(4)	Elk nadeel hep z'n voordeel. Het komt hartstikke goed met klimaatdoelstellingen. Dankzij #coronavirus . CO2 uitstoot in China daalde met maar liefst 25%. De rest van de wereld volgt nu. Greta kan ik haar handjes klappen. https://t.co/wKCLAoxTyy
(5)	Let een beetje op elkaar. Zo komen we er wel. En een dubbel hart onder de riem voor alle verpleegkundigen, mantelzorgers en pgb-hulpverleners die gewoon doorgaan omdat het hun vak en roeping is om te zorgen voor zwakkere kinderen en ouderen. #Corona https://t.co/ASewlQcl7x
(6)	'Zoals het klokje thuis tikt, tikt het nergens.'. Een heerlijk zonnig weekend, wij blijven thuis. Genieten in de tuin, vast weer iets 'opruimen' en lekker veel knuffelen met de meiden. Wat zijn jullie plannen? https://t.co/s19BotVdWR. #blijfthuis #familie #stayhome #corona https://t.co/ZiLPrTPwoG

Table 13: Original examples from the COVID-19 dataset.

The Voice of Holland	**Childcare Benefits Scandal**
#BOOS	#Rutte
John de Mol	VVD
Ali B	@RenskeLeijten
#MarcoBorsato	Nederland
Jeroen Rietbergen	#Ruttedoctrine
Floods	**COVID-19**
#Limburg	Nederland
Maas	RIVM
Duitsland	China
België	Italië
Valkenburg	België

Table 14: Top five most frequent named entities.

Emotion	Retweets	Replies	Likes	Quotes	All
neutral	1.12	1.07	6.93	0.14	9.26
anger	1.17	1.24	8.73	0.18	11.32
fear	0.90	1.45	7.87	0.12	10.34
joy	0.84	1.15	7.55	0.13	9.67
love	2.36	1.80	21.04	0.24	25.44
sadness	0.49	1.13	9.69	0.08	11.38

Table 15: Traffic The Voice of Holland.

Emotion	Retweets	Replies	Likes	Quotes	All
neutral	3.35	1.36	10.08	0.52	15.31
anger	3.55	1.15	10.22	0.44	15.36
fear	2.63	1.40	9.42	0.37	13.83
joy	2.15	1.06	10.41	0.29	13.91
love	3.11	2.26	14.56	0.33	20.26
sadness	3.51	1.87	16.21	0.51	22.08

Table 16: Traffic Childcare benefits scandal.

Emotion	Retweets	Replies	Likes	Quotes	All
neutral	3.55	1.38	10.56	0.59	16.08
anger	2.63	1.63	11.54	0.44	16.24
fear	1.95	1.40	8.43	0.31	12.09
joy	3.19	1.22	17.14	0.38	21.94
love	1.92	1.16	17.97	0.24	21.30
sadness	2.38	1.20	12.85	0.24	16.67

Table 17: Traffic Floods.

Emotion	Retweets	Replies	Likes	Quotes	All
neutral	1.53	0.76	3.73	0.23	6.25
anger	2.48	1.20	7.12	0.31	11.11
fear	1.31	0.88	3.84	0.19	6.22
joy	1.31	0.78	6.80	0.21	9.09
love	1.66	0.77	8.17	0.22	10.83
sadness	2.18	1.55	7.74	0.33	11.80

Table 18: Traffic COVID-19.

0_vrouwen_mannen_seksueel_gedrag (1423)	**40_tim_hofman_debroervanroos_boos (57)**
1_ali_alib_boos_alibouali (1406)	41_vaccinatieplicht_2g_gevaccineerd_vaccineren (57)
2_tvoh_tv_boos_de (1021)	42_slachtoffers_john_mol_de (56)
3_linda_haar_mol_ze (1018)	43_simon_whiskey_tennessee_stem (54)
4_voice_the_zingen_stem (570)	44_itv_talpa_verkocht_rtl (53)
5_marco_borsato_ali_jeroen (343)	45_gate_boos_vivandaag_nou (53)
6_john_mol_hij_de (302)	46_advocaat_advocate_advocaten_beau (49)
7_nederland_nederlanders_bekende (299)	47_rusland_oekraïne_oekraine_oorlog (48)
8_anouk_haar_gele_ik (241)	48_vertrouwenspersoon_leiderschap_je_bedrijf (47)
9_slachtoffers_daders_de_voor (236)	49_macht_machtsmisbruik_misbruik_mensen (46)
10_ik_nooit_mijn_een (232)	50_dubai_ali_in_geldende (46)
11_holland_voice_the_of (225)	51_programma_stoppen_la_generatie (45)
12_prachtig_voila_mooi_haar (211)	52_jefferson_gast_talent_leuke (45)
13_coach_coaches_als_nieuwe (184)	53_youtube_servers_boos_16 (43)
14_nieuws_media_de_is (182)	54_stoelen_stoel_draaistoelen_rode (41)
15_metoo_het_en_van (175)	55_politiek_coronamaatregelen_coronabeleid (40)
16_rtl_rtl4_programma_rtlnieuws (168)	56_artiesten_artiest_theater_die (40)
17_regisseur_naam_nieman_martijn (146)	57_nr_hashtag_afgelopen_trending (39)
18_kandidaten_kandidaat_voor_die (140)	58_corona_variant_nieuws_over (39)
19_twitter_tweet_tweets_ik (127)	59_gooise_matras_pakt_opgeschud (38)
20_oeps_hop_hoeba_jemig (120)	60_boos_even_gijpie_duidelijker (38)
21_marokkaan_knuffel_marokkaanse (118)	61_rtl_the_voice_uitzenden (36)
22_jury_juryleden_jurylid_rechter (110)	62_politie_mbo2_hartvnl_nieuwsuur (35)
23_maan_marco_ze_was (109)	63_racisme_sylvana_racismekaart_huidskleur (34)
24_talentenjacht_talent_talentenjachten (107)	64_grace_glennis_doseren_crême (32)
25_albert_verlinde_humberto_tan (107)	65_kerk_jezus_paus_god (32)
26_bandleider_rietbergen_jeroen_stapt (107)	66_gommers_klaver_klaarmetrutte_paternotte (30)
27_hartvnl_nosjournaal_op1npo_nieuwsuur (90)	67_team_ajax_wereldje_voetbal (29)
28_anouk_corrupte_bende_stopt (87)	68_rietbergen_jeroen_excuses_hij (28)
29_johnny_hlf8_mol_vader (85)	69_familie_mol_lindademol_uppel (26)
30_seksueel_holland_gedrag_of (84)	70_typhoon_maan_en_door (26)
31_kinderen_zoon_je_ouders (83)	71_miljoen_views_kijkers_000 (26)
32_john_johndemol_mol_prins (81)	72_rietbergen_jeroen_dickpic_dubbele (25)
33_grensoverschrijdend_gedrag (72)	73_wie_gewonnen_jamai_jaar (24)
34_glennis_engels_nee_shock (72)	74_trending_topics_vollemaan_ollongren (23)
35_je_ik_gaat_moet (72)	75_roast_eet_eten_misselijk (22)
36_kaart_auditie_prime_doorgestoken (68)	76_molendijk_ronald_flirt_ronaldmolendijk (22)
37_sponsor_mobile_sponsoren_sponsoring (66)	77_cultuur_cultuursector_missionmars21 (21)
38_spotify_podcast_podcasts_apple (63)	78_cdavandaag_schaamjekapotmedia_gemeenteraadsverkiezing (21)
39_reclame_reclames_advertentie_ad (60)	79_glennis_haar_whitney_jotewawe (20)

Table 19: Topics found by BERTopic in The Voice of Holland subset.

0_kabinet_rutte_de_het (12636)	39_2021_2020_jaar_eind (79)
1_nederland_nederlandse_in_nederlanders (3960)	40_niger_kaag_was_sigridkaag (75)
2_ouders_kinderen_gedupeerde_kinderopvangtoeslag (3456)	41_ombudsman_nat_ombudsman_nationale_klachten (73)
3_belastingdienst_de_bij_het (1884)	42_nazi_fascisme_fascisten_fascistische (71)
4_overheid_burgers_regering_de (1269)	43_basisinkomen_inkomen_een_kosten (68)
5_racisme_discriminatie_etnisch_institutioneel (1200)	44_avondklok_weekend_morgen_week (67)
6_politiek_politici_politieke_democratie (1089)	45_vertrouwen_dicteren_herstellen_nieuws (66)
7_media_journalisten_journalistiek_journalist (666)	46_klimaat_klimaatcrisis_stikstofcrisis_crisis (65)
8_cda_omtzigt_hoekstra_het (631)	47_nr_hashtag_trending_uur (61)
9_000_euro_30_geld (594)	48_weekers_uh_pennestreek_getalen (60)
10_debat_over_het_debatten (495)	49_slapen_slaap_wakker_geslapen (57)
11_crimineel_criminelen_criminele_strafrechtelijk (456)	50_lelystadairport_luchtvaart_schiphol_stikstof (55)
12_vaccinatie_vaccinatiestrategie_vaccin_vaccineren (447)	51_palmen_memo_cleyndert_mevrouw (52)
13_schulden_private_schuldeisers_overheid (407)	52_alzheimer_dementie_last_acute (51)
14_fraude_fraudeurs_fraudeur_fraudebestrijding (381)	53_commissie_donner_onderzoekscommissie_van (50)
15_geheugen_herinnering_herinneren_ik (335)	54_leugenaar_liegen_waarheid_leugens (47)
16_ministerraad_notulen_informatie_openbaar (330)	55_machtig_geautomatiseerde_risicoprofielen_rijk (46)
17_rechter_rechters_advocaat_landsadvocaat (287)	56_glas_plas_water_drinken (44)
18_minister_premier_president_hij (278)	57_cartoon_cartoonist_freutel_hugo (43)
19_nr_hashtag_trending_uur (228)	58_onderste_steen_boven_mh17 (41)
20_twitter_tweet_tweets_2ekamertweets (214)	59_sneeuw_ijs_ijsberg_elfstedentocht (39)
21_leiderschap_nieuw_kaag_sigridkaag (188)	60_tv_netflix_resoluut_show (39)
22_syrië_steun_terroristen_syrie (179)	61_crimineel_rutte_strafrechtelijk_vervolgd (39)
23_corona_coronacrisis_coronabeleid_crisis (169)	62_trending_toeslagenafaire_push_hee (39)
24_algoritmes_data_privacy_algoritmen (149)	63_video_filmpje_film_kanaal (39)
25_bulgarenfraude_bulgaren_fraude_marokkanen (144)	64_pensioen_pensioenaffaire_pensioenstelsel_pensioenen (36)
26_kerst_kerstreces_fijne_kerstdagen (143)	65_buikpijn_blokpoel_had_directeur (36)
27_trump_realdonaldtrump_joebiden_minpres (131)	66_catshuisregeling_catshuis_ouders_betaaldatum (36)
28_nr_hashtag_trending_uur (131)	67_fiets_fietsen_fietsje_corfu (36)
29_catshuis_catshuisoverleg_overleg_catshuisregeling (122)	68_je_werk_werknemers_werkgever (35)
30_virus_coronavirus_viruswaarheid_mexicaansegriep (122)	69_belastingdienst_efficiency_zorgbonus_bij (34)
31_migranten_vluchtelingen_migratie_massale (110)	70_hersteloperatie_herstel_herstelorganisatie_uht (34)
32_koning_koningsdag_doneer_republiek (102)	71_doctrine_oprotten_rutte_vrijheid (34)
33_jeugdzorg_zorg_jeugdzorgaffaire_ziekenhuizen (101)	72_euro_000_belastingdienst_30 (33)
34_bananenrepubliek_bananenmonarchie_bananen_een (101)	73_nieuws_goed_hopelijk_eindelijk (33)
35_staatssecretaris_meerwaarde_wiebes_staatsecretaris (93)	74_kerst_doneer_fijne_kinderen (32)
36_armoede_woningnood_arm_verzorgingsstaat (93)	75_hashtag_nr_trending_uur (31)
37_christelijke_christenunie_jezus_christenen (88)	76_club_clubje_clubjes_deze (30)
38_moslims_xenofobische_lastercampagnes_vuilwoorderij (84)	

Table 20: Topics found by BERTopic in Childcare Benefits subset.

0_limburg_de_in_het (2066)	20_2021_2020_juli_peterrdevries (60)
1_klimaatverandering_de_klimaat_het (607)	21_ziekenhuis_venlo_viecuri_patiënten (56)
2_nederland_dat_in_en (244)	**22_ramptoeristen_te_je_mensen (51)**
3_belgië_duitsland_in_en (232)	23_engel_willem_hoax_willemengel (44)
4_dieren_koeien_huisdieren_paarden (190)	24_militairen_leger_60_68 (43)
5_nr_hashtag_trending_nl (144)	25_defensie_ingezet_veiligheidsregio_politie (36)
6_regen_mm_code_neerslag (142)	26_peter_vries_overleden_peterrdevries (36)
7_journalist_telegraaf_nieuws_ramp (132)	27_verzekeraars_verzekeringsmaatschappijen_premie (35)
8_overstromingen_overstroming_inondations_ramp (126)	28_telegraaf_thomas0811_jullie_op (34)
9_boeren_ondernemers_schade_de (120)	29_helikopter_helikopters_vliegen_boven (33)
10_evacuatie_geëvacueerd_valkenburg_evacueren (107)	30_g7_logo_geïntroduceerd_waterlogo (30)
11_koning_griekenland_vakantie_koningin (100)	31_china_zhengzhou_beelden_in (30)
12_l1_radio_uitzending_tv (97)	32_liedje_lied_nporadio2_muziek (30)
13_geld_kabinet_overheid_de (95)	33_drone_dronebeelden_beelden_video (28)
14_vaccinatie_coronavirus_vaccinatieplicht_vaccineren (72)	34_777_giro_opengesteld_geopend (27)
15_slaapplek_een_camping_welkom (71)	35_huizen_bouwen_huis_uiterwaarden (26)
16_00_uur_30_venlo (70)	36_doneren_giro777_help_doneer (24)
17_doneren_giro777_donatie_gedoneerd (69)	37_zwemwaterlocatie_zwemmen_gevaarlijk_let (22)
18_foto_beelden_deze_mooie (64)	38_racisme_donkere_bij1_kleur (21)
19_brandweer_luik_verviers_westhoek (62)	39_je_pcrtest_schokkend_negatieve (21)

Table 21: Topics found by BERTopic in Floods subset.

0_virus_coronavirus_corona_besmettingen (47335)	74_johnson_brexit_boris_britse (879)
1_scholen_kinderen_onderwijs_school (22386)	75_zweden_zweedse_aanpak_stockholm (871)
2_nederland_amsterdam_nederlandse_nederlanders (20864)	76_bloemen_tulpen_bloemetje_planten (864)
3_zorg_ic_patiënten_ziekenhuis (15980)	77_dam_halsema_demonstratie_amsterdam (858)
4_grootse_stopgezet_wereldgezondheidsorganisatie (13161)	78_boeren_landbouw_tuinbouw_boer (856)
5_voetbal_knvb_jpl_sport (10525)	79_strand_stranden_kust_zee (855)
6_china_chinese_wuhan_chinezen (10463)	80_turkije_turkse_erdogan_turken (853)
7_politiek_politici_kabinet_regering (9121)	81_koning_toespraak_alexander_willem (836)
8_belgië_belgische_antwerpen_vlaanderen (7707)	82_demonstratie_demonstranten_protest_demonstreren (820)
9_covid19_covid_19_covid_19 (7223)	83_vrijwilligers_hulp_helpen_vrijwilligerswerk (817)
10_italië_italie_italiaanse_italia (6952)	84_koffie_coffeeshops_coffeeshop_rijen (812)
11_supermarkt_supermarkten_winkels_winkel (6234)	85_banken_bank_rente_rabobank (769)
12_klm_luchtvaart_schiphol_vluchten (6015)	86_belasting_belastingen_belastingdienst_uitstel (758)
13_nieuws_persconferentie_media_journalisten (5669)	87_sekswerkers_seks_sekswerk_prostitutie (729)
14_eu_europa_europese_eurobonds (5605)	88_voedselbank_voedselbanken_voedsel_voedselpakketten (725)
15_mondmaskers_mondmasker_maskers_dragen (5056)	89_sars_cov_sarscov2_virus (725)
16_hashtag_trending_nr_afgelopen (4796)	90_hydroxychloroquine_chloroquine_zink_hcq (721)
17_vaccin_vaccins_vaccinatie_tegen (4660)	91_vrouwen_mannen_vrouwelijke_gender (721)
18_thuiswerken_thuis_werken_werk (4586)	92_humor_lachen_grap_grappen (712)
19_dieren_nertsen_hond_katten (4578)	93_angst_paniek_bang_raadgever (707)
20_coronacrisis_crisis_deze_we (4028)	94_roken_rokers_tabak_sigaretten (705)
21_muziek_lied_concert_zingen (3827)	95_5g_straling_zendmasten_masten (695)
22_trump_realdonaldtrump_president_vs (3712)	96_globalceasefire_wapenstilstand_antonioguterres_moedige (674)
23_app_privacy_apps_apple (3570)	97_cybercriminelen_cybersecurity_phishing_cybercrime (673)
24_kerk_god_kerken_jezus (3478)	98_medicijn_medicijnen_tegen_geneesmiddel (656)
25_klimaat_co2_klimaatverandering_natuur (3104)	99_golf_tweede_2e_2de (636)
26_testen_test_tests_ggd (2953)	100_psychische_mentale_ggz_gezondheid (623)
27_werkloosheid_werknemers_arbeidsmarkt_werkgever (2886)	101_schip_haven_boord_boot (585)
28_ondernemers_ondernemer_mkb_zzp (2688)	102_japan_olympische_spelen_japanse (567)
29_ouderen_oudere_bejaarden_senioren (2456)	103_oorlog_defensie_leger_wereldoorlog (557)
30_lockdown_intelligente_regels_maatregelen (2448)	104_maggie_deblock_maggie_sophie_wilmes_block (527)
31_tv_radio_uitzending_3fm (2342)	105_netflix_netflixnl_series_serie (499)
32_trein_treinen_bus_vervoer (2312)	106_veiligheidsraad_nationale_maatregelen (499)
33_wc_papier_toiletpapier_hamsteren (2224)	107_pensioen_pensioenfondsen_pensioenen_dekkingsgraad (478)
34_spanje_spaanse_madrid_barcelona (2200)	108_bruins_bruno_minister_debat (476)
35_restaurant_restaurants_eten_café (2136)	109_beperking_gehandicaptenzorg_verstandelijke_handicap (470)
36_fietsen_fiets_fietsers_wielrenners (2090)	110_testen_nederland_test_tests (464)
37_twitter_tweet_tweets_ik (2065)	111_verjaardag_jarig_vieren_birthday (463)
38_weekend_dag_feestje_koningsdag (2027)	112_carnaval_gevierd_tilburg_vieren (458)
39_kunst_cultuur_museum_musea (2026)	113_cartoon_heindekort_comic_cartoonoftheday (455)
40_vluchtelingen_asielzoekers_migranten_arbeidsmigranten (1996)	114_india_y2020_jamshedpur_lockdown (455)
41_meter_afstand_houden_anderhalve (1943)	115_quarantaine_quarantine_weken_dagen (452)
42_handen_wassen_handenwassen_zeep (1933)	116_kanker_borstkanker_kankerpatiënten_kankerdiagnoses (425)
43_politie_agenten_agent_politiemensen (1893)	117_rioolwater_water_drinkwater_afvalwater (425)
44_youtube_video_filmpje_film (1866)	118_bubbel_bubbels_veiligheidsraad_10 (413)
45_2020_april_maart_bijeenkomsten (1858)	119_hamsteren_hamsters_hamster_hamsteraars (411)
46_ventilatie_lucht_aerosolen_luchtkwaliteit (1782)	120_festival_festivals_zomer_pinkpop (403)
47_afrika_ebola_afrikaanse_congo (1767)	121_game_spel_spelen_bingo (399)
48_huur_huurders_huurverhoging_woningmarkt (1671)	122_gedetineerden_gevangenis_gevangenen_gevangenissen (397)
49_moslims_ramadan_islam_allah (1652)	123_cijfers_statistieken_statistiek_rivm (397)
50_digitale_online_webinar_digitaal (1627)	124_symptomen_symptoom_ziekte_koorts (393)
51_brazilië_bolsonaro_ecuador_colombia (1606)	125_blog_bloggen_gt_blogs (390)
52_pandemie_pandemic_who_een (1597)	126_viroloog_virologen_vanranstmarc_osterhaus (375)
53_boek_boeken_bibliotheek_gedicht (1573)	127_begrafenis_afscheid_overleden_familie (372)
54_burgemeester_burgemeesters_gemeente_brief (1526)	128_vitamine_vitamines_vitamined_vitaminec (366)
55_alcohol_bier_drinken_wijn (1476)	129_podcast_podcasts_spotify_luister (363)
56_economie_economische_crisis_economisch (1429)	130_complot_complottheorie_complotdenkers (360)
57_auto_verkeer_rijden_snelweg (1370)	131_immuniteit_immuunsysteem_immuun_antistoffen (356)
58_vakantie_camping_zomervakantie_zomer (1312)	132_bruiloft_trouwen_huwelijk_trouwfotograaf (355)
59_iran_iraanse_regime_coronavirus (1216)	133_plasma_sanquin_bloed_bloedgroep (342)
60_wetenschap_wetenschappers_wetenschappelijke (1207)	134_hugo_jonge_cda_hugodejonge (342)
61_racisme_blacklivesmatter_discriminatie_racistisch (1207)	135_open_horeca_deuren_weer (341)
62_marokko_marokkaanse_marokkanen_vliegverbod (1192)	136_advocaat_rechtspraak_rechtbanken_rechtbank (340)
63_rusland_russische_poetin_russen (1158)	137_brandweer_brand_vuur_brandweermannen (340)
64_toerisme_reizen_reis_buitenland (1133)	138_tennis_atp_wta_djokovic (339)
65_griep_influenza_gewone_doden (1128)	139_tuin_blijfthuis_tuincentrum_planten (334)
66_frankrijk_franse_macron_parijs (1054)	140_geld_betalen_terugvraagt_aangedacht (333)
67_brabant_brabantse_noord_provincie (1027)	141_helikopter_medic01_traumahelikopter_mmt (333)
68_f1_formule_gp_formule1 (1004)	142_daklozen_dakloze_dakloos_daklozenopvang (331)
69_israël_israel_israëlische_netanyahu (975)	143_skiën_vindicat_ski_skivakantie (327)
70_zon_lente_wandelen_wandeling (949)	144_tandarts_tandartsen_mondzorg_tandartspraktijk (311)
71_socialdistancing_distancing_social_sociale (947)	145_kleur_kleuren_rood_licht (309)
72_foto_fotograaf_beelden_fotografie (920)	146_vogels_duiven_vogel_natuur (304)
73_korea_zuid_zuidkorea_stijgt (918)	147_exitstrategie_exit_strategie_exitplan (303)

Table 22: Topics found by BERTopic in COVID-19 subset.