International Journal of Communication 17(2023), 5932–5953 1932–8036/20230005

Reading Emotions in the Digital Age: A Deep Learning Approach to Detecting Anxiety During the COVID-19 Pandemic through Social Media

JINWOO JEONG♦
SUJIN YOON
DONGYOUNG SOHN
YONG-SUK CHOI*
Hanyang University, South Korea

During an unprecedented crisis, such as the COVID-19 pandemic, people respond to increased uncertainty in their social surroundings, which often entails various emotional expressions, including fear and frustration. Anxiety is particularly important because it may serve as a symptomatic indicator of various social problems, such as the collapse of trust, polarized public opinions, and an increase in violence. Identifying the expressions of anxiety and tracking their fluctuations over time may lead to a better understanding of how people collectively cope with social crises. This study aims to develop a deep learning-based classification of anxiety and track how the degree of anxiety changed over time in the context of the COVID-19 pandemic. Using the bidirectional encoder representations from transformers (BERT) model, we extracted anxiety-laden messages from Twitter and examined how the longitudinal distribution of anxiety corresponded to the major waves of COVID-19 using intervention time-series analyses.

Keywords: social media, anxiety, COVID-19, deep learning, time-series analysis

Anxiety, an unpleasant psychological response to perceived threats or uncertain circumstances, plays a pivotal role in societal functioning. Heightened anxiety in a population can trigger a cascade of social problems, from eroding trust and polarizing public opinion to escalating rates of violence and suicide (Nepon, Belik, Bolton, & Sareen, 2010; Valentino, Banks, Hutchings, & Davis, 2009). As anxiety does not merely affect individual well-being but can influence the social fabric and collective decision-making processes,

---

Jinwoo Jeong: jinwoohg@hanyang.ac.kr
Sujin Yoon: geeksayo@gmail.com
Dongyoung Sohn: dysohn@hanyang.ac.kr
Yong Suk Choi: cys@hanyang.ac.kr
Date submitted: 2022-07-29

This work was supported by the National Research Foundation of Korea (NRF) Grant (No. 2018R1A5A7059549, No. 2020R1A2C1014037) and by Institute of Information & Communications Technology Planning & Evaluation (IITP) Grant (No. 2020-0-01373), funded by the Korea government(*MSIT). *Ministry of Science and Information & Communication Technology

understanding and quantifying the prevalence and impact of anxiety is crucial for effective social and policy interventions, including protecting vulnerable populations, effectively communicating policies, and managing crises (Lee & Kwon, 2021).

Amid unparalleled crises, the significance of quantifying and scrutinizing anxiety levels across various regions and overtime dramatically increases. During crises such as the COVID-19 pandemic, for example, individuals habitually seek updates on the situation, tracking news sources to mitigate uncertainty and perceived threats. These routines, however, may inadvertently amplify anxiety because of the dynamic and often inconsistent nature of the information presented through the news media (Lee, 2020; Liu, Bartz, & Duke, 2016). Furthermore, the digital interconnectedness of modern society can make individuals more sensitive to the actions of their neighbors, more exposed to unpleasant emotions and experiences, and more susceptible to misinformation related to the crisis at hand, such as COVID-19 (Hou, Bi, Jiao, Luo, & Song, 2020).

This suggests that the current media environment does more than just facilitate the dissemination of both accurate and misleading news; it also amplifies the transmission of negative emotions, such as anxiety, which in turn influences how individuals engage with their social surroundings. During the pandemic, for instance, it was reported that nearly half of the South Korean population experienced a surge in depression and anxiety, a phenomenon referred to as "Corona Blue" (Son & Heo, 2020). These issues highlight the pressing need for effective methods to identify and monitor societal distress and anxiety in real time. However, robust methods for accurately monitoring social distress and anxiety remain largely undeveloped, creating a significant challenge in pinpointing which regions or societies necessitate preventive interventions.

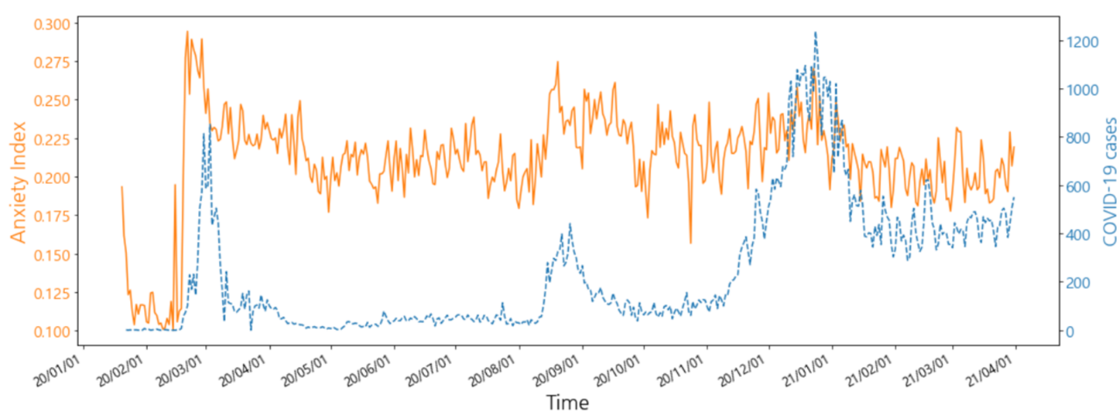

Figure 1. Time series of the confirmed COVID-19 cases[1](dashed line) and daily anxiety index (solid line) extracted and measured from the messages posted on Twitter in South Korea.
Note. The figure was created by the authors.

[1] The number of confirmed COVID-19 cases in South Korea is available at https://corona-live.com/

Social media provides a rich tapestry of human emotion, manifested through both verbal methods, such as text, and nonverbal expressions like emoticons and photos. Textual messages are particularly useful for emotion identification and the study of emotional reactions to specific situations (Gaind, Syal, & Padgalwar, 2019). Although traditional emotion-measuring techniques rely primarily on questionnaires, innovative approaches are now using the vast amounts of data shared on social media to detect sentiments and emotions unobtrusively. For example, a recent study pioneered the development of a machine learnin model capable of real-time detection of anxiety-laden messages on Twitter, using this to create a geographicagl anxiety map (Choi, Kim, & Sohn, 2020). However, this promising approach had limitations, notably the lack of longitudinal anxiety pattern examination and the insufficient sophistication of the Naïve Bayes classification model used for optimal accuracy.

Our study aims to address these issues by constructing a more advanced and accurate anxiety classifier, leveraging the power of deep learning and the bidirectional encoder representations from transformers (BERT) model. With this refined tool, the objective of this study is not only to identify and monitor anxiety levels across time but also to verify the statistical correlation between detected anxiety fluctuations and the patterns of COVID-19 spread, shown in Figure 1, using intervention time series. By connecting the dots between anxiety trends and the course of the pandemic, we hope to paint a clearer picture of the impact unprecedented crises can have on societal anxiety and beyond.

Social Media Users as Human Social Sensors

Human behavior is deeply intertwined with social relationships (Tarde, 1903), with people's thoughts and actions often depending on the behaviors of others, both near and far. The likelihood of voting in an election or getting vaccinated largely depends on how many others in one's social circle exhibit similar behaviors. Instances of increased suicides following the news of celebrity suicides or the popularity of suicidogenic songs, such as "Gloomy Sunday" (Rezső, 1933), further illustrates the tendency for people to conform to others' behaviors, even extreme ones with irreversible consequences like suicide (Lester & Gunn, 2011; Mueller & Abrutyn, 2015). Although numerous explanations have been proposed for such social influence processes (Cialdini & Goldstein, 2004), the role of human emotions remains largely unknown and indeterminate.

People may react emotionally to uncertain situations or threats, potentially transmitting their emotional states to others nearby. Earlier thinkers like Le Bon (1895) identified emotional contagion as a key process through which group members lose their individuality and assimilate with others in large crowds. Although complete assimilation is unlikely, human sociality depends heavily on individuals' abilities to determine what other people feel and do (Galesic et al., 2021). Emotional contagion plays a crucial role in interpersonal communication through which individuals' facial expressions, voices, and attitudes are coordinated (Hatfield, Cacioppo, & Rapson, 1993). Crisis situations can amplify emotional contagion, as uncertain individuals are more likely to observe others' behaviors, accelerating the sharing and diffusion of emotions like anxiety and fear. In such situations, collective behaviors largely depend on individuals' attentiveness to one another and what types of emotions widely shared in the group (Horwitz, 2010; Wang, Zhang, Lin, Zhao, & Hu, 2016).

Anxiety can intensify when individuals face uncertain futures or outcomes in threatening or difficult situations and feel unable to cope or overcome them (Sarason, Sarason, & Pierce, 1990). This may be exacerbated when individuals experiencing anxiety are exposed to others in similar situations, especially on social media. Studies have found that individuals highly sensitive to uncertainty tend to react strongly to crises (e.g., the spread of the H1N1 virus) and other people's responses to them, which worsens their anxiety (Taha, Matheson, & Anisman, 2014). This process can eventually affect even moderately sensitive individuals, potentially leading to societal consequences. Social media platforms expand the scope of observing others' feelings and behaviors, possibly accelerating the diffusion and societywide increase of emotions like anxiety (Primack et al., 2017).

To understand and address the spread of negative emotions such as anxiety, a systematic method for detecting early warning signals is necessary. Shared anxious feelings can lead to severe individual or societal responses, such as depression, conflicts, violence, and even suicide (Wang et al., 2016). Tracking emotional responses can help prevent extreme behaviors and allow for a multilevel perspective on human behaviors based on collective emotions, extending beyond individual sentiments. However, most previous studies have treated anxiety as an individual-level problem requiring therapeutic treatment (Choi et al., 2020) and primarily employed self-report measures or indirect methods of identifying past behavior through recall or tracing.

Although memory-based approaches can be useful for various purposes, they struggle to capture the volatile nature of individual sentiments (Chung & Zeng, 2020; Dasborough, Sinclair, Russell-Bennett, & Tombs, 2008; Scherer, 2005). Longitudinal surveys, for example, find it difficult to examine the subtle changes in sentiments overtime, and privacy concerns or social desirability bias may also cause individuals to withhold their true feelings or provide misleading answers. To overcome these limitations, we propose an alternative observational approach that considers individual social media users as *human social sensors* (Galesic et al., 2021), monitoring their social surroundings on a much-finer timescale. Social media platforms enable users to express their emotions in real time, creating a rich data source that reflects the emotional climate of communities. In the face of unpredictable situations, individuals tend to become more attentive to others' responses to uncertain events, which may be influenced by their positions in a socially networked environment (Boursier, Gioia, Musetti, & Schimmenti, 2020; Lwowski, Rad, & Choo, 2018).

Natural language expressions on social media can be seen as immediate responses of human social sensors to uncertain local or global environments. These expressions often provide unfiltered emotional information that traditional methods might not capture. By monitoring communication on social media and extracting various emotion-laden messages using supervised machine learning (SML), we can obtain real-time observations of how numerous individuals distributed across time and space perceive and experience various events in a society (Weiler, Grossniklaus, & Scholl, 2016). Working with social media users as effective human social sensors of emotional spread can enable researchers to better understand the dynamics of emotional contagion and its impact on collective behaviors, especially in times of crisis.

This observational approach is not new, and attempts have been made to identify sentiments by analyzing textual messages on social media for various purposes. Most of these efforts focused on identifying valence, with relatively few attempts to detect specific types of emotions, such as anxiety. Some pioneering examples used Bayesian SML models to detect anxiety levels from large corpora in social media and analyzed the relationships between anxiety and various social phenomena (e.g., Choi et al., 2020; Gruda & Hasan, 2019). Although the Bayesian SML approach has been widely applied and proven useful for various tasks and problems, there is a growing interest in applying diverse, deep learning models to natural language processing, including recurrent neural networks (Zhao, Lu, Cai, He, & Zhuang, 2017), convolutional neural networks (Dos Santos & Gatti, 2014), memory networks (Hazarika et al., 2018), and attention mechanisms (Majumder et al., 2019).

In this work, we attempt to develop an advanced anxiety classifier using BERT, as it has demonstrated better performance in text classification tasks, including sentiment analysis, than many other SML models (Devlin, Chang, Lee, & Toutanova, 2019). By taking advantage of BERT's capabilities, we aim to create a more accurate and efficient anxiety detection system that can analyze large-scale social media data in real time. This study primarily focuses on the following research questions:

RQ1: Will the anxiety classifier developed in this study outperform existing classifiers in effectively detecting anxiety levels in social media content?

RQ2: Does the anxiety classifier developed in this study generate valid classification results reflecting important social events, such that the detected anxiety levels indeed reflect the patterns of COVID-19 spread in different time periods?

By addressing these research questions, we hope to contribute to a better understanding of collective emotions during times of crisis and demonstrate the potential of using social media as a valuable data source for analyzing public emotions in real time. In the following sections, we describe the model developed to extract anxiety-laden tweets and examine how these messages are associated with the confirmation of COVID-19 cases.

Methods

Detecting Anxiety in Social Media

BERT is a deep learning model designed to represent text semantics. It consists of self-attention layers with learnable parameters that change their values during model training. During the training phase, the parameters in each self-attention layer are adjusted to assign larger weights to the important spans of characters (tokens) in a sentence (Devlin et al., 2019).

BERT is pretrained using corpora such as English Wikipedia in advance of training for any given task, such as sentiment classification. During pretraining, a random token in a sentence is masked, and the model attempts to predict the correct word of the masked token using the unmasked tokens in the sentence. Furthermore, the model receives a pair of consecutive sentences as input and learns to predict whether the second sentence in the pair is a subsequent sentence in the original document (Devlin et al., 2019).

The BERT version used in this study was KorBERT.[2] KorBERT is a BERT for Korean language, which was presented by the Electronics and Telecommunications Research Institute in Korea. It employs the same architecture and pretraining method as BERT. However, pretraining and tokenization methods for dividing sentences into tokens are tailored to capture the unique characteristics of the Korean language. The corpus used for pretraining KorBERT is Korean reading materials, such as Korean newspaper articles and encyclopedias, and the tokenization method separates each sentence into Korean morphemes, the smallest semantic units. KorBERT was then fine-tuned to develop our BERT-based anxiety classifier. In this context, fine-tuning refers to the training process for a specific task, such as sentiment classification, after pretraining.

Originally, when fine-tuning the BERT-based classifier presented by Devlin et al. (2019), sentences are separated into tokens, and the [CLS] special token for representing the sentence-level classification context is inserted at the beginning of the sentence. Then, all the tokens are fed into the pretrained BERT, which returns the hidden representation of each token in vector form, and the hidden representation of the [CLS] token summarizes the semantic context of the sentence. Subsequently, the hidden representation of the [CLS] token is used as input for the feed-forward layer. When an input vector (hidden representation) is fed to the feed-forward layer, the final scalar output is returned by calculating the matrix multiplication between the input vector and learnable weight vectors (Goodfellow, Bengio, & Courville, 2016). Finally, the model is trained to minimize the cross-entropy loss between the returned output and correct class value. The cross-entropy loss measures the performance of a classification model (Goodfellow et al., 2016). Using this loss, the model can gradually minimize the difference between the correct class value and returned output during the training phase. The cross-entropy loss is calculated as follows:

$$Cross\ Entropy\ Loss = -\frac{1}{N}\sum_{j=1}^{N} y_j * \log \hat{y}_j\,, \tag{1}$$

where N is the number of sentences, j is the index of a sentence, y_j is the correct class value of sentence j, and $\hat{y}_j$ is the returned class output of sentence j.

In our model, instead of simply using the [CLS] token's hidden representation fed into the feed-forward layer for classification, we used every token's hidden representation, except for that of the [CLS] token. This model was designed to capture the meaning of the entire sentence by considering the contextual information provided by all the tokens. By incorporating the hidden representations of all tokens, the model can better understand and analyze the overall semantic context of a given sentence (Reimers & Gurevych, 2019).

---

[2] KorBERT is available at http://aiopen.etri.re.kr/service_dataset.php. In the fine-tuning phase, the dimension of token's hidden representation was 768, the learning rate was $5e^{-5}$, and the batch size was 8 with 2 epochs. All the other hyperparameters were set to default of KorBERT.

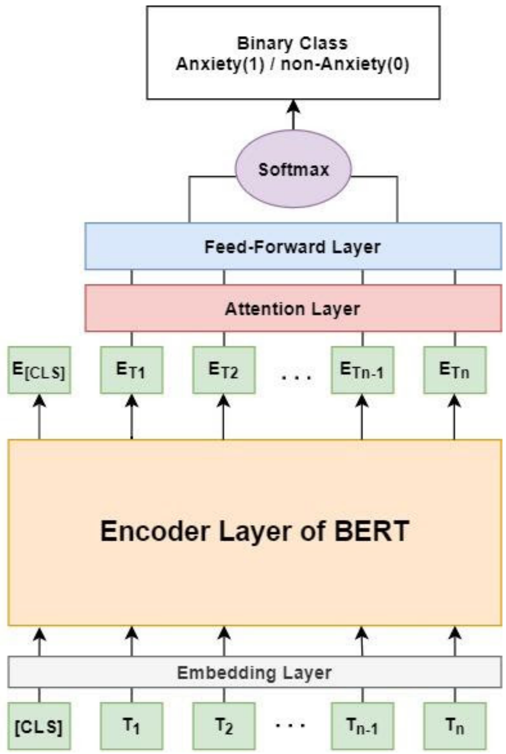

Figure 2. Architecture of our BERT-based anxiety classifier.

Note. The figure above was created by the authors.

Furthermore, all the hidden representations of the tokens were fed into an attention layer, a feed-forward layer, and finally a softmax layer, which returned the probability of each class as a scalar output (Goodfellow et al., 2016).

First, in the attention layer, the alpha score of each token was calculated. Given a sentence $S \in R^{k \times d}$ of k length (the number of tokens) and d dimension of S_i (hidden representation of i-th token in sentence S) and a learnable weight vector $w \in R^d$, the alpha score α_i of the i-th token was calculated as follows:

$$\alpha_i = \frac{\exp(S_i \cdot w)}{\sum_{j=1}^{k} \exp(S_j \cdot w)}. \tag{2}$$

Importantly, this alpha score was designed as the weight of each token, which indicates the influence of each token on the classification result. The alpha score was multiplied by the hidden representation of each corresponding token and then summed. This process assigned more weight to more important tokens when classifying input sentences and produced more effective sentence representations (Vaswani et al., 2017). The output of this attention layer was fed to the feed-forward and softmax layers.

This process is illustrated in Figure 2, where T_n represents the *n*-th input token, and E_{Tn} denotes the hidden representation of token T_n.

To construct a data set to fine-tune our BERT-based anxiety classifier, tweets were collected from February 2016 to June 2020, using an Open API Tweet crawler called "*tweepy*" (Roesslein, 2020), and 78,022 messages were then randomly chosen from the crawled tweets and tagged as anxiety-laden or not. In the process, seven graduate students majoring in communications at a large metropolitan university in Korea were recruited to serve as human coders. Since anxiety is an elusive emotion, often manifested in the guise of discomfort, anger, or frustration, the coders were instructed to consider explicit expressions of anxiety as well as others symptomatic of the state of "being anxious," particularly in relation to uncertain situations, including environmental threats and unknown events. Before tagging 78,022 tweets, the coders tagged 800 tweets, 1% of the sample, as either anxiety-related or not anxiety-related and the intercoder agreement measured using the Holsti score turned out 92.67%.

Among 78,022 tagged tweets, 70% were randomly used for training (fine-tuning), 15% for validation, and 15% for testing. The training data set consisted of data used to train the model. The validation data set was used to evaluate and adjust the model. Because the model was adjusted based on its performance on the validation data set, it could have been overfitted to the validation data set. Therefore, a test data set was required to evaluate the final tuned model.

Table 1. Number of Tweets in the Data Set.

	Training (before duplication)	Validation	Test
Anxiety	49,650 (3310)	709	710
Nonanxiety	51,305	10,994	10,994

Note. Data collected from Twitter by authors.

Table 2. Results of Anxiety Classifiers Tested on the Test Data Set.

Model	Macro F1-score	Accuracy
Naïve Bayes Classifier	0.5878	0.8432
Original BERT-based Classifier	0.6446	0.8761
Our BERT-based Anxiety Classifier	**0.6658**	**0.9082**

Table 3. Confusion Matrix of Our BERT-Based Anxiety Classifier Tested on the Test Data Set.

		Predicted value		*Total actual value*
Actual value		*Anxiety*	*Nonanxiety*	
	Anxiety	*331 (0.47)*	*379 (0.53)*	*710*
	Nonanxiety	*696 (0.06)*	*10298 (0.94)*	*10994*

Note. Values in the parentheses indicate ratios with the total actual value as the denominator.

The ratio of tweets tagged as "non–anxiety-related" to "anxiety-related" was 14.4:1. To avoid problems of biased classification resulting from data imbalance (Thabtah, Hammoud, Kamalov, & Gonsalves, 2020), anxiety-related tweets were duplicated up to 15 times in the training data set only (Table 1). Table 2 shows the results of our fine-tuned BERT-based anxiety classifier tested on the test data set, along with those of the Naïve Bayes classifier (Choi et al., 2020) and the original BERT-based classifier, which uses only the hidden representation of the [CLS] token for classification (Devlin et al., 2019). The results suggest that our model outperformed the other two models in accuracy and Macro F1-score. This implies that our model detects the classes (anxiety or nonanxiety) more accurately and that the predicted classes are less biased toward one specific class. For more detailed results, the confusion matrix that shows the result of applying our model to the test data set is presented in Table 3. Using our model, inferences were elicited from the tweets collected for analysis, and the anxiety label of each tweet was extracted.

To analyze the level of anxiety during the COVID-19 outbreak, we collected tweets posted from January 20, 2020, to March 31, 2021. The starting date corresponds to the emergence of the first COVID-19 case in South Korea. After data collection, duplicate tweets posted by the same user each day were eliminated to avoid selection bias. Subsequently, 2,249,045 tweets were selected and classified using our BERT-based anxiety classifier, resulting in 474,324 tweets being classified as anxious tweets. Given that the total number of days was 437, approximately 5,147 tweets were posted on average per day, of which approximately 1,085 were anxious tweets. After classifying all tweets, an anxiety index, indicating the level of anxiety, was calculated as the ratio of anxious tweets to all the tweets posted during a specific period. To compare the anxiety index between different time periods, it was calculated over specific durations (range of dates), such as the time during a major wave. Furthermore, for time-series analysis, the daily anxiety index was calculated from January 20, 2020, to March 31, 2021.

Results

Descriptive Analysis of the Anxiety-Related Social Issues

Specifying all reasons people feel anxious is extremely cumbersome, but we attempted to make inference indirectly by examining what issues seem related to anxiety-laden tweets. Accordingly, we calculated the document ratio, which refers to the percentage of documents containing a particular term (Robertson, 2004). By treating each anxious tweet as a document and each term as a noun that represents a certain event or issue, the document ratio can show which events or issues most triggered the posting of anxiety-laden tweets.

Using a Korean morpheme analyzer *KoNLPy* (Park & Cho, 2014), all nouns from anxiety-related tweets were extracted. Then, from the extracted nouns, those with the top-five document ratio values were selected, as listed in Table 4. "COVID-19" was the word with the highest document ratio during the entire study period, followed by words related to politics: "Democrats, Prosecution, President," and "Lee Jae Myung," who is a former governor of Gyeonggi-do, the largest province in South Korea.

Furthermore, we investigated how the relevant issues changed overtime by dividing the entire period into three major waves of COVID-19 (February 19 to April 19, 2020; August 15 to October 11, 2020; and November 17, 2020, to February 14, 2021). "COVID-19" had the highest document ratio, with document ratios of 0.10, 0.07, and 0.04, in the respective waves.

Table 4. Nouns With Top-Five Document Ratios in the Entire Study Period.

Noun	Document Ratio
COVID-19	0.046
Democrats	0.020
Prosecution	0.018
President	0.017
Lee Jae Myung[a]	0.008

Note. [a]Lee Jae Myung is a former Korean governor of Gyeonggi Province. Data collected from Twitter by the authors.

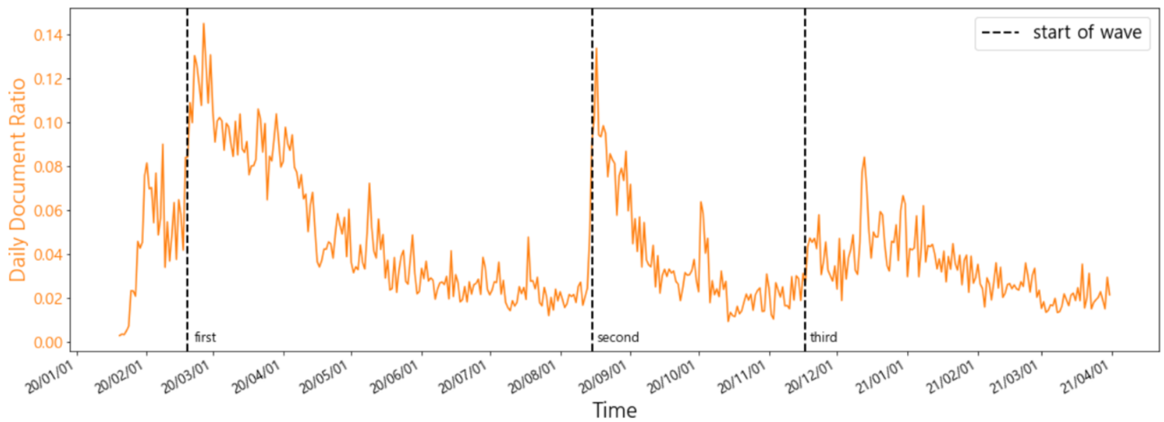

Figure 3. Daily document ratio of the Korean word for "COVID-19" and the beginnings of major waves.

Note. Data collected from Twitter by the authors.

On the dates of the beginnings of the three major waves (February 19, 2020; August 15, 2020; and November 17, 2020), 1547, 1299, and 1085 anxiety-laden tweets, respectively, were posted. Among them, the word with the highest document ratio was "COVID-19," which appeared 130, 118, and 28 times, and the corresponding document ratios were 0.088, 0.091, and 0.026, respectively, as shown in Figure 3. Furthermore, by comparing the document ratio of a particular word on a certain date with those on other dates, we could examine whether that the word had an abnormal effect on anxiety levels on that particular date. The top four most frequently appearing words on each date were then selected. None of the selected nouns were related to COVID-19, and each noun represented a certain issue. Table 5 shows the document ratios of the selected nouns on the designated dates. Then, using the daily document ratio of each selected word, a box plot was constructed to check if the document ratio of a selected noun on a certain date was higher than usual.

Figure 4 shows the distributions of the daily document ratios of the selected words. When the document ratios of the nouns in Table 5 are matched and compared with the box plots of the document ratios of the nouns in Figure 4, none of the document ratios in Table 5 are found to be outliers. This implies that the issues unrelated to COVID-19 did not have an unusual effect on the anxiety level on the days when the major COVID-19 waves started.

Table 5. Nouns With Top 4 Document Ratios That Are Unrelated to COVID-19 on the Start Dates of Major COVID-19 Waves.

Start date of the first wave		Start date of the second wave		Start date of the third wave	
Noun	Document Ratio	Noun	Document Ratio	Noun	Document Ratio
Democrats	0.035	President	0.015	Prosecution	0.022
President	0.020	Politics	0.014	Yoon Seok Yeol[b]	0.018
Moon Jae In[a]	0.016	Moon Jae In[a]	0.014	Press	0.016
Politics	0.014	Pro-Japanese	0.013	Politics	0.015

Note. [a]Moon Jae In is the former president of South Korea. [b]Yoon Seok Yeol is the current president of South Korea. Data collected from Twitter by the authors.

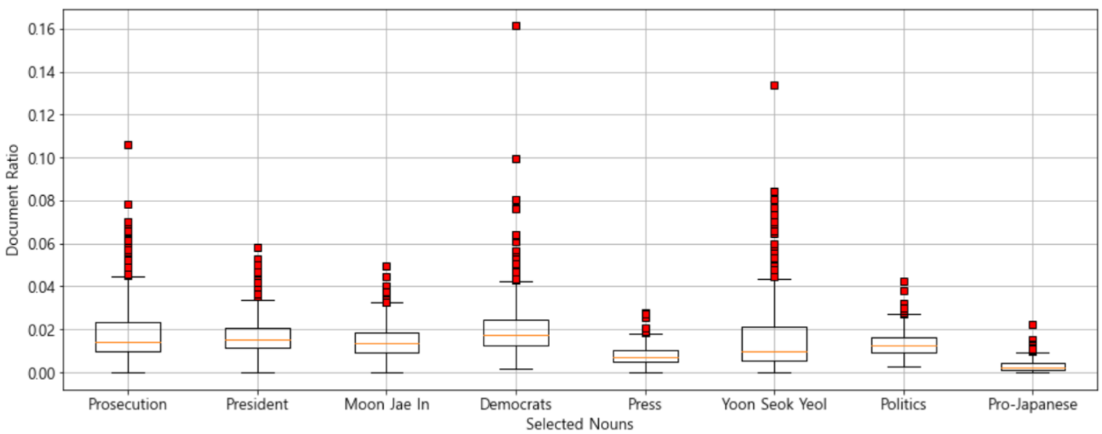

Figure 4. Box plot of document ratios of selected nouns during the entire study period.

Note. The horizontal line of the bottom whisker indicates the minimum value, bottom horizontal line of the box refers to the first quartile, horizontal line in the middle of the box indicates the median, top horizontal line of the box is the third quartile, and horizontal line of the top whisker is the maximum. Furthermore, the dots indicate outliers. Data collected from Twitter by the authors.

Table 6. Anxiety Index in Different Time Periods.

Time Period	Anxiety Index (Total number of tweets)
Entire study period	0.2109 (2,249,045)
Before the first wave	0.1186 (185,410)
During the first wave	0.2373 (407,266)
Between the first and second waves	0.2072 (627,685)
During the second wave	0.2313 (297,711)
Between the second and third waves	0.2161 (168,525)
During the third wave	0.2197 (387,706)
After the third wave	0.2009 (174,742)

Note. Data collected from Twitter by the authors.

Anxiety Fluctuations During Major Waves of the Pandemic

To see an overall trend of the anxiety level, anxiety indices for seven meaningful time periods were compared. The time periods were divided based on the starting and ending points of each major wave of COVID-19. Table 6 summarizes the anxiety indices with seven subperiods. A cursory look at the results revealed that the anxiety indices were higher during the three major COVID-19 waves than during the other periods and higher during the first wave than during later waves. The anxiety index was lowest before the first major wave, indicating that the COVID-19 outbreak was the major cause of the anxiety spike.

Because the anxiety index was calculated for a range of dates, the results indicate different anxiety levels in different time periods, which can be easily interpreted. However, such an approach ignores the time-series aspect of our data set (e.g., autocorrelation and trend), which is gathered sequentially overtime. Therefore, to analyze the changing pattern of the anxiety index accurately, the daily anxiety index was calculated, and an intervention time-series analysis, with major waves of COVID-19 as interventions, was conducted.

We conducted a segmented regression analysis, which is often used for easy interpretation of intervention effects with time-series data (Schaffer, Dobbins, & Pearson, 2021; Wagner, Soumerai, Zhang, & Ross-Degnan, 2002). The aim was to examine how the daily anxiety index changed according to each of the three COVID-19 waves. The equation for segmented regression is as follows:

$$Y_t = \alpha + \beta_1 \times time_t + \beta_2 \times intervention_t + \beta_3 \times time\_since\_intervention_t + \epsilon_t \quad (3)$$

where Y_t is the dependent variable at an instant *t*, and $time_t$ is the number of days since the start of the study period at a time point *t*. $Intervention_t$ is a dummy variable that indicates whether instant *t* is before (0) or after (1) the implementation of the intervention, and $time\_since\_intervention_t$ represents the number of days elapsed since intervention implementation until instant *t*, taking a value of 0 before the intervention. Parameter α denotes the intercept, which implies the baseline level, β_1 represents the baseline trend, β_2 indicates the level change during the intervention (increased or decreased amount), β_3 is the trend change during the intervention, and ϵ_t is an error term at instant *t*.

To encode the major wave of COVID-19 as an intervention in the model ($intervention_t$), a step function was employed, which could encode an intervention for a certain time period as follows:

$$intervention_t = \begin{cases} 0, & if\ t < T \\ 1, & otherwise \end{cases} \quad (t = time, T = beginning\ of\ major\ wave) (4)$$

Before fitting the model with the daily anxiety index, we divided the study period into three subperiods to capture and model the different time-series aspects, such as the difference in trend and autocorrelation of the daily anxiety index between the major COVID-19 waves, as shown in Table 7.

Table 7. Time Period Description for Segmented Regression.

Time Period	Start Date	Beginning of Intervention	End of Intervention (End Date)
1	Jan. 20, 2020 First COVID-19 case in South Korea	Feb. 19, 2020 Start of the first wave	Apr. 19, 2020 End of the first wave
2	Apr. 20, 2020 The day after the end of the first wave	Aug. 15, 2020 Start of the second wave	Oct. 11, 2020 End of the second wave
3	Oct. 12, 2020 The day after the end of the second wave	Nov. 17, 2020 Start of the third wave	Feb. 14, 2021 End of the third wave

Note. Dates are announced by Korean Disease Control and Prevention Agency (2021).

Before conducting a time-series analysis, it is important to check the stationarity and autocorrelation problems of the daily anxiety index. To do so, an augmented Dickey-Fuller test to test the stationarity of the time-series data set was performed. The *p*-values of the test in all three time periods were below 0.05, showing significant stationarity. However, the Durbin-Watson statistics for the residual of each segmented regression model at lag 1 were 1.325, 1.296, and 1.261, in order of time periods, indicating the presence of autocorrelation. Failing to correct autocorrelation may lead to underestimated standard errors and overestimated significance of the effects of an intervention. Therefore, the AR(1) error term, which represents the autoregressive error at lag 1, was added to the segmented regression equation (Huitema, Mckean, & Mcknight, 1999; Kim, Kim, & Lee, 2020; Penfold & Zhang, 2013). The results of the Durbin-Watson test on the residuals of segmented regressions with AR(1) error in the order of the time period were 2.140, 2.059, and 1.928, significantly reducing the autocorrelations of the residuals.

Tables 8, 9, and 10 show the results of the segmented regression with the AR(1) error for each time period. The baseline trends ($\beta_1^1,\ \beta_1^2,\ \beta_1^3$) were statistically insignificant, indicating that no systematic patterns were found across times. The results also showed that the level of anxiety increased during the major waves of COVID-19 when compared with the period before ($\beta_2^1 > 0$, $\beta_2^2 > 0$, $\beta_2^3 > 0$), and the magnitude of change was the greatest in the first wave and then decreased ($\beta_2^1 > \beta_2^2 > \beta_2^3$). This implies that the initially spiked anxiety level, mainly because of the first major wave, diminished as the later waves came. In addition, although the results show a negative trend during the first and third waves ($\beta_3^1,\ \beta_3^3$), the difference was not statistically significant. However, during the second wave, the index gradually decreased overtime ($\beta_3^2 < 0$). In summary, the results imply the possibility of the existence of psychological desensitization.

Table 8. Results of Segmented Regression With AR(1) Error in Period 1 (January 20, 2020, to April 19, 2020).

	Coefficient	Std error	*t*-statistic	Log-Likelihood
Baseline level α^1	0.0806^{***}	0.011	7.015	
Baseline trend β_1^1	−0.0001	0.000	−0.412	244.304
Change in level during first wave β_2^1	0.1001^{***}	0.012	8.049	
Change in trend during first wave β_3^1	−0.0004	0.000	−1.097	

Note. The superscripts of α and β refer to the time period corresponding to Table 7, and the subscript is the specific coefficient of a variable that corresponds to Equation (3).
$^*p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

Table 9. Results of Segmented Regression With AR(1) Error in Period 2 (April 20, 2020, to October 11, 2020).

	Coefficient	Std error	*t*-statistic	Log-Likelihood
Baseline level α^2	0.1214^{***}	0.014	8.599	
Baseline trend β_1^2	0.00001	0.00004	0.327	509.966
Change in level during second wave β_2^2	0.0255^{*}	0.005	5.204	
Change in trend during second wave β_3^2	-0.0004^{***}	0.000	−3.806	

$^*p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

Table 10. Results of Segmented Regression With AR(1) Error in Period 3 (October 12, 2020, to February 14, 2021).

	Coefficient	Std error	*t*-statistic	Log-Likelihood
Baseline level α^3	0.1308^{***}	0.019	7.087	
Baseline trend β_1^3	−0.00002	0.000	−0.086	344.171
Change in level during third wave β_2^3	0.0122^{*}	0.007	2.213	
Change in trend during third wave β_3^3	−0.0002	0.000	−0.910	

$^*p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

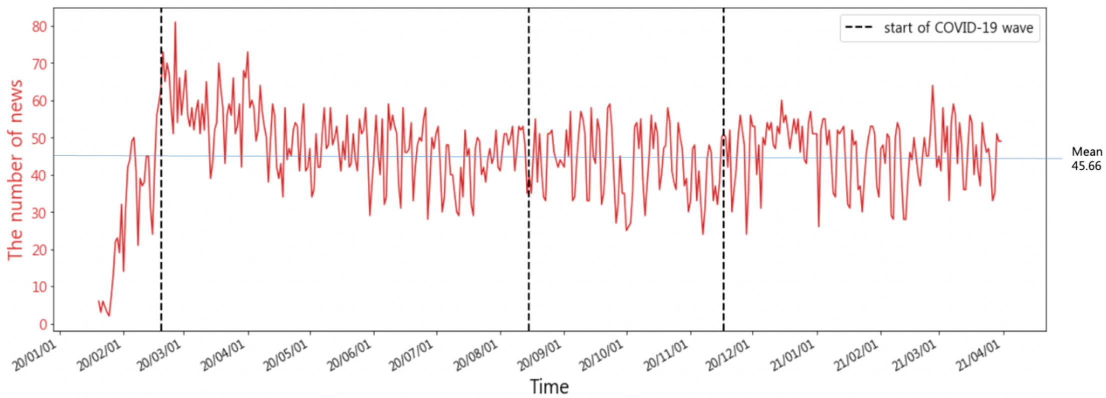

Figure 5. The number of news reports about COVID-19.

Note. The Y-axis represents the number of news articles uploaded online. Data collected from *BIG KINDS*[3] by the authors.

One might suggest the declining trends of anxiety as resulting is not from the psychological desensitization processes but mostly from the mere quantitative decease of COVID-19 related news overtime—the less news about the disease is circulated, the less anxious people become about the disease. To see if this might be the case, we further examined how the number of relevant news reports changed during the same period in Korea. "BIG KINDS" (BIG KINDS, 2022) is a widely used news big data analysis system in South Korea for news media research (Kim, 2019). Using this system, we collected news reports related to COVID-19 published by five major South Korean daily newspapers (*Kyunghyang*, *Chosun*, *Joongang*, *Donga*, *Hankyoreh*). The news articles were categorized under disease and health and filtered using keywords such as "COVID-19, Corona virus, COVID," and "Corona," in both English and Korean. As a result, we gathered a total of 19,954 news reports for analysis. As shown in Figure 5, except for the days when the COVID-19 outbreak started, the number of news coverage turned out remaining relatively stable overtime with no apparent trend of declining from the mean (45.66). This implies that, given the relatively stable news coverage of the disease, it seems reasonable to attribute the declining trends of anxiety found to some other causes, including psychological desensitization.

Discussion

Our statistical analysis has highlighted the efficacy of the advanced classifier we developed in pinpointing messages laden with anxiety. Remarkably, these identified patterns seem to mirror the real-world timeline of the pandemic, offering a distinct real-time insight into collective societal anxiety. As the initial number of confirmed COVID-19 cases started to rise, we noted an indication of heightened anxiety, likely driven by the fear of high infection risk. Yet, as the pandemic endured, there was a perceptible shift in people's psychological responses; they appeared to acclimatize to the pandemic conditions, leading to a

[3] BIG KINDS is operated by the Korea Press Foundation (KPF), a prestigious nonprofit organization established in 1961 to promote the development of the nation's journalism and media industry. BIG KINDS collects and archives news articles from major newspapers, magazines, and broadcast news transcripts.

gradual desensitization of their reactions. This was particularly evident when we observed that the emergence of the first two infection waves induced a significant surge in anxiety, while this pattern did not hold during the third wave of the pandemic.

The prospect of such desensitization implies that, as the duration of the pandemic extends, people might become less responsive overtime to disease control and public health regulations/policies. This evolving public sentiment poses additional challenges to pandemic management. The power of real-time monitoring of societal emotional trends is particularly noteworthy here, as it may provide government agencies with the ability to detect early shifts in public coping strategies. Consequently, this can facilitate the development of alternative approaches and policies that are more in sync with the changing emotional landscape.

Although the findings of our study are based on the COVID-19 pandemic and Korean society, they can be applied and extended to broader issues and situations. Monitoring real-time communication can help identify early signals of social anomalies of any kind, which can be critical in various scenarios beyond pandemics. The ability to detect and analyze emotional patterns on social media platforms presents a valuable tool for policy makers, researchers, and practitioners alike in addressing a wide range of social concerns. For example, the real-time monitoring of social media communication could be applied to detect early signs of public opinion polarization, civic unrest, or environmental disasters. By identifying these early signals, government agencies can take proactive measures to mitigate potential negative consequences, develop more effective policies, and allocate resources more efficiently. Moreover, real-time monitoring can facilitate the evaluation of existing policies and interventions, enabling decision makers to make necessary adjustments in response to the changing social landscape.

Beyond the temporal monitoring of public emotions, the findings of this study can be extended to the possibility of monitoring and mapping the spatial distribution of emotions like anxiety across a nation, which enables the identification of specific geographic regions with potential problems (Choi et al., 2020). By analyzing the spatial patterns of collective emotions, researchers and policy makers can gain a deeper understanding of regional differences and potential sources of social stress. Consequently, tailored interventions can be developed to address the unique needs of these regions, leading to more efficient allocation of resources and targeted policy measures. Additionally, the advanced anxiety classifier developed in this study can be adapted to analyze different emotions and sentiments, such as fear, anger, or happiness, which could be relevant in various contexts, like natural disasters, elections, or economic crises. This versatility allows for a more comprehensive understanding of societal trends, collective emotions, and their potential impacts on various aspects of social life.

As social media platforms continue to reshape the way people communicate and interact with one another, understanding the dynamics of emotional contagions and the role of social media in spreading emotions becomes increasingly important in communication research. Probing real-time communication on social media allows researchers to explore how the public emotionally responds to various stimuli, such as established news agendas as well as the burgeoning issues of fake news (O'Connor, Balasubramanyan, Routledge, & Smith, 2010). It invites exploration into whether certain types of emotions drive the rapid spread of misinformation or rumors, how social networks influence emotional states, and the potential repercussions of these phenomena on both individual and societal well-being (Coviello et al., 2014; Kramer, Guillory, & Hancock, 2014). Understanding the intricate tapestry of emotional transmission and spread in the digital age offers insights into the impact of our interconnected world on emotional well-being.

Furthermore, our research showcases the benefits of leveraging advanced machine learning techniques, such as BERT-based classifiers, for sentiment analysis. Such advanced tools pave the way for creating more precise and intricate models that can provide a more comprehensive understanding of the emotional landscape in various contexts. As the field of communication research undergoes continuous evolution, the incorporation of these emerging techniques and methodologies will allow researchers to uncover new insights and contribute to a deeper understanding of the complex interplay between emotions, communication, and social dynamics.

Limitation

This study has some limitations. First, it relies solely on data from Twitter, which may result in platform-dependent results. Active Twitter users are not representative of general social media users or the wider population, as they tend to be more politically active than users of other platforms like Instagram. Additionally, individual differences in expressing anxiety on social media may be influenced by personality traits, such as extraversion (Gruda & Ojo, 2022), which could lead to over or underrepresentation of certain voices on Twitter. Moreover, user behaviors on social media may be significantly shaped by platform-specific algorithmic features (Malik & Pfeffer, 2016). To achieve more generalizable results, future research should examine multiple channels of communication, including search engines and various social media platforms, to determine whether similar patterns emerge across different platforms.

Second, the model used in this study classifies anxiety as a binary variable. This approach was chosen because of the difficulty in identifying the intensity of anxiety from short Twitter texts (up to 140 characters) and concerns that labeling anxiety as a continuous variable may increase human bias. Additionally, using multiple anxiety classes can lower the accuracy of the SML model (Gruda & Hasan, 2019). However, even short texts may convey varying degrees of anxiety. Further research should explore methods for establishing clear thresholds for anxiety intensity within short texts and develop a model that can accurately classify multiple levels of anxiety.

Third, anxiety can sometimes be masked by other emotions, such as anger, frustration, or fatigue. To ensure the discriminant validity of the classification results, it is important to investigate the relationships between anxiety and other relevant emotions. Additionally, communicated messages may have multiple meanings, with text-based messages and atypical data carrying various intended denotations. Noise in these messages can make it challenging to accurately identify specific emotions. For example, an anxiety-laden tweet may contain a mix of information about COVID-19 anxiety and other personal circumstances. Future research should aim to develop a model that accurately measures anxiety by distinguishing between different situations, such as crises, public or social events, and personal circumstances.

Conclusion

Anxiety can proliferate rapidly and extensively in a networked communication environment, where individuals are continually exposed to the emotional responses of others. Amid pervasive uncertainties such as the COVID-19 pandemic, a collective perception that the future is unpredictable can escalate individual anxiety into a shared societal concern. Monitoring the expression of anxiety on social media platforms allows for real-time tracking and early detection of significant issues, which can help prevent a situation from spiraling further. To develop an effective real-time anxiety monitoring system, our study was structured around two primary research questions. The results for RQ1 demonstrated the efficacy of our anxiety classifier, which surpassed existing models in accurately detecting anxiety levels in social media content. The accuracy and F1-score of our model led to its use in real-time monitoring of anxiety levels in South Korea during the pandemic. As for RQ2, we found that the classifier developed here generated valid results reflecting key social events, aligning detected anxiety levels with patterns of COVID-19 spread in different periods. The statistical analysis confirmed that the anxiety index developed here captured temporal shifts in anxiety corresponding to disease waves, as well as the individual tendency toward desensitization to environmental uncertainty overtime.

References

BIG KINDS. (2022, October 17). *News search and analysis: COVID-19*. Retrieved from www.bigkinds.or.kr/v2/news/search.do

Boursier, V., Gioia, F., Musetti, A., & Schimmenti, A. (2020). Facing loneliness and anxiety during the COVID-19 isolation: The role of excessive social media use in a sample of Italian adults. *Frontiers in Psychiatry, 11*. doi:10.3389/fpsyt.2020.586222

Choi, Y. S., Kim, H., & Sohn, D. (2020). Mapping social distress: A computational approach to spatiotemporal distribution of anxiety. *Social Science Computer Review, 40*(3), 598–617. doi:10.1177/0894439320914505

Chung, W., & Zeng, D. (2020). Dissecting emotion and user influence in social media communities: An interaction modeling approach. *Information and Management, 57*(1), 103–108. doi:10.1016/j.im.2018.09.008

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*, 591–621. doi:10.1146/annurev.psych.55.090902.142015

Coviello, L., Sohn, Y., Kramer, A. D., Marlow, C., Franceschetti, M., Christakis, N. A., & Fowler, J. H. (2014). Detecting emotional contagion in massive social networks. *PLoS One, 9*(3), 1–6. doi:10.1371/journal.pone.0090315

Dasborough, M. T., Sinclair, M., Russell-Bennett, R., & Tombs, A. (2008). Measuring emotion: Methodological issues and alternatives. In N. M. Ashkanasy & C. L. Cooper (Eds.), *Research companion to emotions in organizations* (pp. 197–208). Cheltenham, UK: Edwin Elgar Publishing (New Horizons in Management Series).

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, 1*, 4171–4186. doi:10.18653/v1/N19-1423

Dos Santos, C. N., & Gatti, M. (2014). *Deep convolutional neural networks for sentiment analysis for short texts*. Paper presented at International Conference on Computational Linguistics, Dublin, Ireland.

Gaind, B., Syal, V., & Padgalwar, S. (2019). *Emotion detection and analysis on social media*. arXiv preprint. Retrieved from https://arxiv.org/abs/1901.08458

Galesic, M., de Bruin, W. B., Dalege, J., Feld, S., Kreuter, F., Olsson, H., . . . van der Does, T. (2021). Human social sensing is an untapped resource for computational social science. *Nature, 595*, 214–222. doi:10.1038/s41586-021-03649-2

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep feedforward networks. In S. Lefrançois (Ed.), *Deep learning* (pp. 161–220). Cambridge, MA: MIT Press.

Gruda, D., & Hasan, S. (2019). Feeling anxious? Perceiving anxiety in tweets using machine learning. *Computers in Human Behavior, 98*, 245–255. doi:10.1016/j.chb.2019.04.020

Gruda, D., & Ojo, A. (2022). All about that trait: Examining extraversion and state anxiety during the SARS-CoV-2 pandemic using a machine learning approach. *Personality and Individual Differences, 188*, 1–6. doi:10.1016/j.paid.2021.111461

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Science, 2*(3), 96–99. doi:10.1111/1467-8721.ep10770953

Hazarika, D., Poria, S., Zadeh, A., Cambria, E., Morency, L., & Zimmermann R. (2018). Conversational memory network for emotion recognition in dyadic dialogue videos. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1*, 2122–2132. doi:10.18653/v1/N18-1193

Horwitz, A. V. (2010). How an age of anxiety became an age of depression. *The Milbank Quarterly, 88*(1), 112–138. doi:10.1111/j.1468-0009.2010.00591.x

Hou, F., Bi, F., Jiao, R., Luo, D., & Song, K. (2020). Gender differences of depression and anxiety among social media users during the COVID-19 outbreak in China: A cross-sectional study. *BMC Public Health, 20*(1), 1–11. doi:10.1186/s12889-020-09738-7

Huitema, B. E., Mckean, J. W., & Mcknight, S. (1999). Autocorrelation effects on least-squares intervention analysis of short time series. *Educational and Psychological Measurement, 59*(5), 767–786. doi:10.1177/00131649921970134

Kim, J. H. (2019). An exploratory study of health inequality discourse using Korean newspaper articles: A topic modeling approach. *Journal of Preventive Medicine and Public Health, 52*(6), 384–392. doi:10.3961/jpmph.19.221

Kim, S. Y., Kim, H., & Lee, J. T. (2020). Health Effects of air-quality regulations in Seoul Metropolitan area: Applying synthetic control method to controlled-interrupted time-series analysis. *Atmosphere, 11*(8), 868–872. doi:10.3390/atmos11080868

Korean Disease Control and Prevention Agency. (2021, July 8). *COIVD-19 cases*. Retrieved from https://ncov.kdca.go.kr/lastBannerList.do?brdId=3&brdGubun=39

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, 111*(24), 8788–8790. doi:10.1073/pnas.1320040111

Le Bon, G. (1895). *The crowd: A study of the popular mind*. London, UK: Trans-Action.

Lee, J.-E., & Kwon, S. (2021). A study on the public's crisis management efficacy and anxiety in a pandemic situation-focusing on the COVID-19 pandemic in South Korea. *Sustainability, 13*(15), 83–93. doi:10.3390/su13158393

Lee, S. A. (2020). Coronavirus Anxiety Scale: A brief mental health screener for COVID-19 related anxiety. *Death Studies, 44*(7), 393–401. doi:10.1080/07481187.2020.1748481

Lester, D., & Gunn, J. F. III. (2011). National anthems and suicide rates. *Psychological Reports, 108*(1), 43–44. doi:10.2466/12.PR0.108.1.43-44

Liu, B. F., Bartz, L., & Duke, N. (2016). Communicating crisis uncertainty: A review of the knowledge gaps. *Public Relations Review, 42*(3), 479–487. doi:10.1016/j.pubrev.2016.03.003

Lwowski, B., Rad, P., & Choo, K. R. (2018). Geospatial event detection by grouping emotion contagion in social media. *IEEE Transactions on Big Data, 6*(1), 159–170. doi:10.1109/TBDATA.2018.2876405

Majumder, N., Poria, S., Hazarika, D., Mihalcea, R., Gelbukh, A., & Cambria, E. (2019). DialogueRNN: An attentive RNN for emotion detection in conversations. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(1), 6818–6825. doi:10.1609/aaai.v33i01.33016818

Malik, M., & Pfeffer, J. (2016). Identifying platform effects in social media data. *Proceedings of the Tenth International AAAI Conference on Web and Social Media, 10*(1), 241–249. doi:10.1609/icwsm.v10i1.14756

Mueller, A. S., & Abrutyn, S. (2015). Suicidal disclosures among friends: Using social network data to understand suicide contagion. *Journal of Health and Social Behavior, 56*(1), 131–148. doi:10.1177/0022146514568793

Nepon, J., Belik, S. L., Bolton, J., & Sareen, J. (2010). The relationship between anxiety disorders and suicide attempts: Findings from the National Epidemiologic Survey on Alcohol and Related Conditions. *Depression and Anxiety, 27*(9), 791–798. doi:10.1002/da.20674

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *Proceedings of the International AAAI Conference on Web and Social Media, 4*(1), 122–129.

Park, E. L., & Cho, S. (2014). *KoNLPy: Korean natural language processing in Python*. Paper presented at the Annual Conference on Human and Language Technology, Chuncheon, Gangwon, South Korea.

Penfold, R. B., & Zhang, F. (2013). Use of interrupted time series analysis in evaluating health care quality improvements. *Academic Pediatrics, 13*(6 Suppl.), S38–S44. doi:10.1016/j.acap.2013.08.002

Primack, B. A., Shensa, A., Escobar-Viera, C. G., Barrett, E., Sidani, J. E., Colditx, J. B., & James, A. E. (2017). Use of multiple social media platforms and symptoms of depression and anxiety: A nationally-representative study among U.S. young adults. *Computers in Human Behavior, 69*, 1–9. doi:10.1016/j.chb.2016.11.013

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embedding using Siamese BERT Networks. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (pp. 3982–3992). Hongkong, China: Association for Computational Linguistics. doi:10.18653/v1/D19-1410

Rezső, S. (1933). *Gloomy Sunday* (Song; Video file). Retrieved from https://youtu.be/9dZj7YW5oFQ

Robertson, S. (2004). Understanding inverse document frequency: On theoretical arguments for IDF. *Journal of Documentation, 60*(5), 503–520. doi:10.1108/00220410410560582

Roesslein, J. (2020). *Tweepy: Twitter for python!* (Source code). Retrieved from https://github.com/tweepy/tweepy

Sarason, B. R., Sarason, I. G., & Pierce, G. R. (Eds.). (1990). *Social support: An interactional view*. Hoboken, NJ: John Wiley & Sons.

Schaffer, A. L., Dobbins, T. A., & Pearson, S. A. (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: A guide for evaluating large-scale health interventions. *BMC Medical Research Methodology, 21*(58), 1–12. doi:10.1186/s12874-021-01235-8

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*(4), 695–729. doi:10.1177/0539018405058216

Son, Y. J., & Heo, M. S. (2020). A study on social media usage, helplessness, and loneliness experienced by college students since the COVID-19 pandemic. *Journal of Digital Contents Society, 21*(11), 1957–1971. doi:10.9728/dcs.2020.21.11.1957

Taha, S. A., Matheson, K., & Anisman, H. (2014). H1N1 was not all that scary: Uncertainty and stressor appraisals predict anxiety related to a coming viral threat. *Stress and Health, 30*(2), 149–157. doi:10.1002/smi.2505

Tarde, G. (1903). *The laws of imitation* (E. C. Parsons & F. H. Giddings, Trans.). New York, NY: Holt.

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences, 513*, 429–441. doi:10.1016/j.ins.2019.11.004

Valentino, N. A., Banks, A. J., Hutchings, V. L., & Davis, A. K. (2009). Selective exposure in the Internet age: The interaction between anxiety and information utility. *Political Psychology, 30*(4), 591–613. doi:10.1111/j.1467-9221.2009.00716.x

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 6000–6010.

Wagner, A. K., Soumerai, S. B., Zhang, F., & Ross-Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics, 27*(4), 299–309. doi:10.1046/j.1365-2710.2002.00430.x

Wang, X., Zhang, L., Lin, Y., Zhao, Y., & Hu, X. (2016). Computational models and optimal control strategies for emotion contagion in the human population in emergencies. *Knowledge-Based Systems, 109*, 35–47. doi:10.1016/j.knosys.2016.06.022

Weiler, A., Grossniklaus, M., & Scholl, M. H. (2016). Situation monitoring of urban areas using social media data streams. *Information Systems, 57*, 129–141. doi:10.1016/j.is.2015.09.004

Zhao, Z., Lu, H., Cai, D., He, X., & Zhuang, Y. (2017). Microblog sentiment classification via recurrent random walk network learning. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence* (pp. 3532–3538). Melbourne, Australia: AAAI Press. doi:10.24963/ijcai.2017/494