

Text Based Mood Classifier

Team 10

21700554 Eungaang Lee

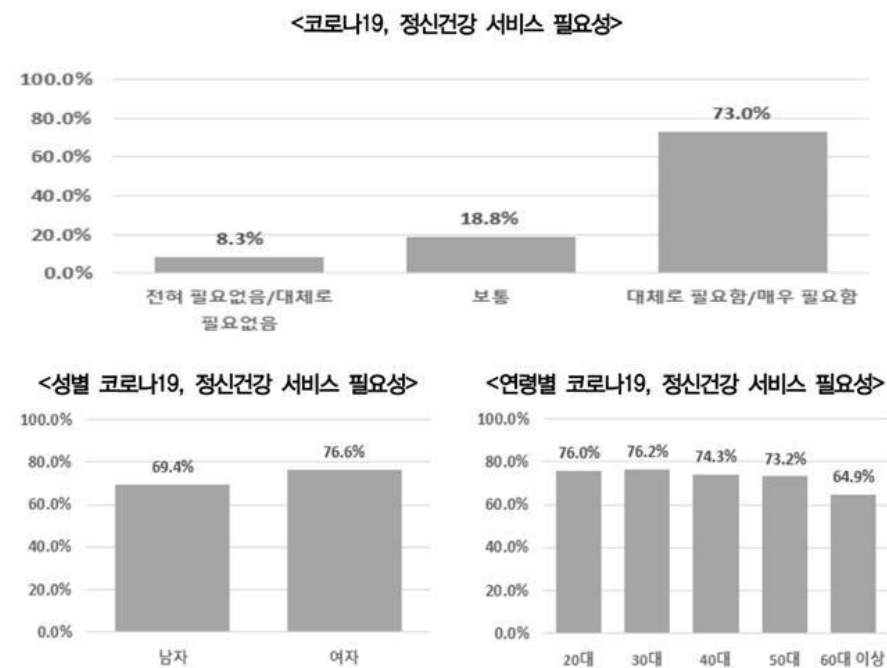
21800088 Seo-ye Kim

21800545 Yewon Lee

21900421 SeongHyeon An

1. Problem Description

We have currently been suffering from COVID-19 epidemic for a year and a half, and because of its impact, many people are experiencing mental health difficulties from mild to severe. Even the new word ‘Corona blue’ appeared, which refers to depression caused by COVID-19.



자료 : 경기연구원(2021), 코로나19로 인한 국민 정신건강 실태조사.

According to a survey conducted by the Gyeonggi Research Institute (2021), more than half of the people feel uneasy and depressed in Korea due to the prolonged threats of COVID-19. The above graph shows that many people feel the need for mental health care services more after COVID-19. We focus on this need through this project.

We decided to set a big goal for the project as to manage people's emotional states using AI techniques. Specifically, we aimed to create a model that can detect or identify people's emotional states from some data. As for the data, we targeted text data because we have a lot of text data inside our smart devices, such as KakaoTalk messages, Twitter, ect.

2. Background and Conventional Approach

There are already many attempts to analyze emotional states from text. There are also many services that use these technologies. As an example, Naver has applied emotion analysis technology that can detect emotions from text to shopping services since 2014, helping anyone to conveniently check, analyze, and utilize reviews.

In addition, there are also attempts to use this technology with similar objectives as ours. The paper "*Analysis and Recognition of Depressive Emotion through NLP and Machine Learning* (2020)" proposes a machine learning based emotional analysis system that detects users' depression through the contents of articles posted on SNS. In this study, the authors used the GetOldTweet3 package to collect data from twitter, and used NLP techniques and the wordCloud package for preprocessing and filtering text data. As machine learning algorithms, they used RNN(Recurrent Neural Network), LSTM(Long Short-Term Memory), and GRU(Gate Recurrent Unit) models. All of these are implemented based on Python 3.6.

As we can see in these studies, and as we have learned in class, we generally follow a process of data collection, data processing using NLP techniques, and prediction using machine learning algorithms in order to predict emotions from text data.

3. Proposed Method

A) Data collection and preparation

In order to build such a classifier, a decent amount of text message training data that is classified into different emotions is needed. The data we were able to acquire on kaggle contained 3579 distinct twitter messages that are classified into one of the four emotions joy, fear, anger and sadness. The image below shows a small portion of the data collected.

1	text	label
2	Just got back from seeing @GaryDelaney in Burslem. AMAZING!! Face still	joy
3	Oh dear an evening of absolute hilarity I don't think I have laughed so mu	joy
4	Been waiting all week for this game ?△톨?△톨?△톨 #cheer #friday ?△톨	joy
5	@gardiner_love : Thank you so much, Gloria! You're so sweet, and though	joy
6	I feel so blessed to work with the family that I nanny for ?△톨 nothing bu	joy
7	Today I reached 1000 subscribers on YT!! , #goodday, #thankful	joy
8	@Singaholic121 Good morning, love! Happy first day of fall. Let's make so	joy
9	#BridgetJonesBaby is the best thing I've seen in ages! So funny, I've misse	joy
10	Just got back from seeing @GaryDelaney in Burslem. AMAZING!! Face still	joy
11	@IndyMN I thought the holidays could not get any more cheerful, and th	joy
12	I'm just still . So happy .WnA blast	joy
13	It's meant to be!! #happy #happy	joy
14	?뽕?뽕톨Yeah?속톨 PAUL?속톨?뽕톨?뽕 #glorious #BB18	joy
15	My morning started off amazing!! Hopefully the whole day is going as i wa	joy
16	?샐 @cailamarsai you've had me ?뽕 뽕 the whole time watching @black_	joy
17	@iamTinaDatta love you so much #smile ?뽕뽕?뽕뽕	joy
18	@WyoWiseGuy @LivingVertical however, REI did offer me the job today as	joy

We also need additional data to be used when testing the model. We were able to acquire 3091 text messages classified into four emotions, joy, fear, anger and sadness that do not overlap with the training data.

B) AI approach

Two main AI techniques, natural language processing(NLP) and random forest were used to build the classification model. The NLP stage is subdivided into 5 tasks. The first task is to convert all uppercase letters into lowercase. This is to properly count the words. The task is to remove handles. Certain characters inside a text message can lead to false counting of the words. Remove handles is a process that prevents such behavior by deleting such characters beforehand. The third task in NLP is to tokenize emojis. Emojis inside text messages contain crucial information about a user's emotion. In order to take these into account, this certain stage tokenizes such emojis in order to consider them as valid tokens. The fourth stage is to remove stopwords. Stopwords are common words that are used too frequently that have low value in modeling a prediction model. Removing stopwords enhances performance and compute time, The final step in NLP is to vectorize the words. This step enumerates all valid words and counts how many each word was used.

Based on the vectorized words obtained by NLP, we constructed a Random forest to generate the mood classification model. The random forest can be seen as an extension of the decision tree method. The random forest model maintains multiple decision trees rather than one for better accuracy

C) Evaluation criteria

The evaluation criteria we set up was getting an overall accuracy over than 0.8 on the test data, and to accurately classify text messages with emojis.

4. Results

The inputs for testing the AI model and the results are shown below. The AI model can classify four emotions such as anger, fear, joy, and sadness. Four emotions were successfully predicted according to the texts. All the input texts are famous quotes and real texts from social networking services.

Input	Result
Weak people take revenge. Strong people forgive. Intelligent people ignore.	Anger
First College Math Test tomorrow.	Fear
This is a beautiful day that the Lord has made. I will rejoice and be glad :)	Joy
This has been the most depressing week full of rain ever lol.	Sadness

The accuracy and overall result of our AI model was analyzed by the heatmap. The heatmap is shown below. The heatmap was calculated with our AI model and dataset for testing which is already provided by Kaggle. The overall accuracy level of our AI model is around 0.856 which is good enough to successfully predict most cases. The accuracy levels of four emotions anger, fear, joy, and sadness are around 0.821, 0.955, 0.848, and 0.757, respectively. Among the four emotions, fear has the highest accuracy level. It means that emotion fear has

the highest data density in both trained and tested datasets.

Accuracy on validation: 0.8559077809798271



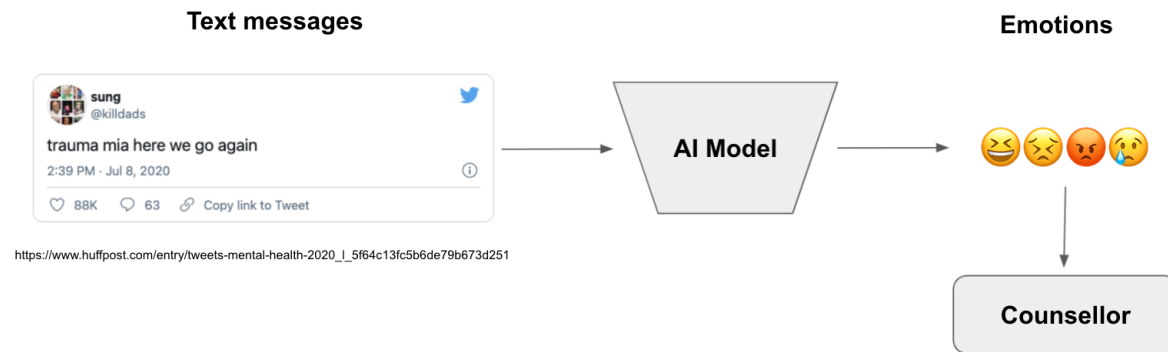
We also tested the texts with emojis to verify whether the emojiification phase was successfully applied. The result is shown below. The general emoji cases were successfully predicted by the AI model. However, if the input emojis were not tokenized during the emojiification phase, the results were concluded as fear.

Input	Result
:)	Joy
):	Fear
XD	Fear

There are many texts which are hard to predict. The typical example is “Hello.” We tested “Hello” as the input of the AI model and the result was fear. The emotion fear has the highest data density and distribution along with the trained and tested datasets, and it led the emotion fear to have the highest accuracy level compared to the other emotions. When the trained AI model cannot predict the texts, it returns the fear as a failure all the time.

5. Conclusion

Expected effect: Record the user's emotional state and utilize it



We were able to classify text messages by people into anger, fear, joy, and sadness to construct a prediction model using NLP and a Random Forest algorithm.

Even though our results include some failures such as return fear when the model cannot classify text messages, it could be used for detecting depressing messages.

We expect that these attempts will help to develop application programs that can record a person's emotional changes and utilize it. For example, if the app detects any dangerous signals from the user, such as depression or suicide, it may be able to connect him/her to a counseling center, of course, based on a clear agreement of the user.

6. Division of Works

All of our group members participated in the process, ranging from selecting project topics to finding datasets, implementing AI models, and organizing the process.

In particular, in the case of video presentation and final report, we divided the parts as below.

Eungaang Lee: Proposed method

Seo-ye Kim: Problem description

Yewon Lee: Conclusion

SeongHyeon An: Results

7. Discussion

Eunganng Lee

This project helped me read and understand python code. Before doing the project, I had some difficulties coding in python, but this project helped my python coding skills. I also thought that knowing all the specific details and techniques of AI may not be that important. Many libraries and codes exist that support those needs. It is more important to have creative ideas on what you want to do with AI and helping the society.

Seo-ye Kim

To be honest, I was not very interested in AI until I took this class, because it felt very ambiguous and boring. However, I was able to actually apply AI techniques to a specific goal through this project. In the process, I could understand that AI is not just about importing and applying existing models. It was interesting that there are processes of processing data and finding out the best model for better results. Thus, one regret about this project is that the models I've learned were limited, so I couldn't try the process of finding the best model. Above all, it was most enjoyable to think about how to apply AI technology to help people.

Yewon Lee

I really appreciate that I was able to participate in implementing an AI model through this project and learn a lot about AI. In the future, I think I will remember what I learned at this time and be more interested in AI.

SeongHyeon An

I realized that developing an AI model is difficult as it has many steps and calculations. Among the several steps, I think that crawling and processing the data are the most important steps. According to datasets, the performance of the AI model is far different. At first, I had difficulties in learning and following the concepts of AI. However, by this term project, I could enjoy and learn about AI and how it worked. I believe that AI is a powerful tool to change the world. I would like to focus and learn more about AI.

References

- Dataset : <https://www.kaggle.com/anjaneyatripathi/emotion-classification-nlp>
- [기사] 코로나 19 로 우울감 1 년 전보다 8.3% 증가 ... 국민 56% 불안, 우울
(홍용덕 기자) : <https://www.hani.co.kr/arti/area/capital/991804.html>
- 감정을 분석하다 - 네이버 감정 분석 AI
https://m.blog.naver.com/naver_search/222080132277
- *Analysis and Recognition of Depressive Emotion through NLP and Machine Learning*
(2020) : <https://www.koreascience.or.kr/article/JAKO202015762902196.pdf>