

자연어 처리와 SNS를 활용한 우울감 초기 디텍터

2014312743 서수민

1. 서론

성인이 된 이후로 많은 친구들이 슬퍼하고, 우울해하고, 무기력해 하는 것을 목격했다. 시간이 지날수록 아픈 사람들이 줄어들거나 상태가 호전되기는 커녕, 무거운 사회적 기대와 스스로에 대한 엄격한 태도로 병원에조차 선뜻 발들이지 못한 채, 혼자 짊어지고 가는 모습을 보았다. 그리고 많은 친구들이 글이나 사진을 창구로 그들의 우울함을, 절박함을 드러내기도 했다. 인스타그램이나 페이스북, 트위터는 읽는 사람도, 쓰는 사람도 가볍게 생각하는 매체이기 때문이다. 그렇게 작은 징조들이 그들의 주변사람들을 무심히 스쳐가고 순식간에 눈덩이로 불어나 아픈 사람들은 압도당했다.

1) 버지니아 공대 총기난사 사건

어렸을 적 미국 유학길에 올랐던 한국인 유학생은 학창시절 극심한 왕따와 인종차별에 시달려 우울증을 갖게 되었다. 그는 친구나 교수님이 말을 걸어도 전혀 대답하지 않았고, 오직 글로서 자신의 분노를 드러냈다. 그가 영문학 수업 시간에 제출한 작문 과제 <Richard McBeef>은 증오로 가득 차 있었고, 재창작되고 문학적으로 분석되기도 하였다. 그러한 증오를 발단으로 그는 끔찍한 총기난사 사건을 일으켰고, 그가 정신과 약을 복용했던 것 때문에 정신질환과 사회 안전에 대한 문제가 제기되기도 했다 (연합뉴스, 2007) (김기성, 2007).

2) 유명 가수의 자살

작년 12월, 국내 인기 가수였던 종현이 우울증을 앓다가 자살로 생을 마감했다 (선미경, 2017). 그가 작사, 작곡한 노래들은 진정성 있는 가사들로 청자에게 위로를 전해주었

다. 고인이 생전 마지막으로 게시한 글에는 디어 클라우드의 노래 '네 곁에 있어'의 캡처 화면과 노래 가사인 "네가 아파하지 않길 기도해 단지 네가 행복하기를 바래"라는 내용이 담겨 있었다 (김경옥, 2018). 많은 사랑을 받던 가수의 사망에 대중은 크게 충격을 받았고, 일부 팬들은 우울감을 호소했다 (대중문화팀스포츠한국, 2017).

3) 기존 연구

트위터, 인스타그램 등 SNS 게시글로 감정을 분석했던 기존 연구들이 있다. 한 연구는 인스타그램의 해시태그를 이용해, 행복과 우울의 감정에 따른 이미지 분석 (사진 색조, 사진 객체 등)을 연구하여 정확성 높은 결과를 얻었다 (김기홍한경식, 2018). 또 다른 연구는 트윗의 감정 분석을 통해 감정을 추출하여, 음악추천 시스템을 개발했는데, SNS 기반의 음악추천에 따른 만족도가 대조군에 비해 높은 것을 확인하였다 (최홍구황인준, 2012). 이처럼 SNS 게시글에서 감정을 추출할 수 있다면, 심리상담, 상품 추천 등으로 응용될 가능성이 충분하다.

이 프로젝트의 목표는 1) SNS 크롤링을 통해 SNS 게시글에 드러난 우울증 초기 증상을 파악해, 2) 나아가 사용자에게 적절한 조치를 추천해주고, 3) '마음을 읽어주는 일기장' 또는 음성비서 서비스에 적용할 수 있는 자연어 처리 프로그램을 만드는 것이다.

2. 실험방법

1) 인스타그램 크롤링¹

GitHub의 Python으로 작성된 인스타그램 크롤링 API²를 사용했다. 인스타그램은 보안

¹ 웹 크롤러(web crawler)는 조직적, 자동화된 방법으로 월드 와이드 웹을 탐색하는 컴퓨터 프로그램이다. 자동 이메일 수집과 같은 웹 페이지의 특정 형태의 정보를 수집하는 데도 사용된다. (위키백과)

² <https://github.com/iammrhelo/InstagramCrawler>

상의 이유로 XPath 경로가 업데이트 마다 수정되기 때문에, HTML 상 첫 포스트 위치, 버튼 위치(XPath) 등의 오류를 모두 수정하고 사용하였다. 다음을 커맨드 창에서 실행하면 Firefox를 실행하여 Geckodriver를 통해 Firefox를 원격제어 하고, 게시글을 긁어 텍스트 파일로 저장한다.

```
instagramcrawler.py -q #happy -n 100 -c
```

실행에 성공하면 다음의 화면이 나온다.

```
...#iamhappy#affirmation#happiness#optimisticquotes#positivequotes#gratitude#lawof attraction#iamabundance#begood']  
Saving...  
Saving to directory: ./data/happy.hashtag  
Quitting driver...
```

인스타그램에서 크롤링한 우울감 관련 해시태그는 #depression (600개), #depressed (100개) 등 총 **700**개의 게시글을 모았다. 행복 관련 해시태그는 #happy (400개), #happiness (200개), #iamhappy (300개) 등 총 **900**개의 게시글을 모았다.

2) 트위터 크롤링³

파이썬 프로그램에서 *tweepy* 모듈을 활용했다. 트위터 API를 사용하려면 Consumer key, consumer secret key, access token, access secret token 등의 네 가지 접근 권한을 얻어야 한다⁴. *tweepy*를 이용하여 액세스를 요청하고 명령을 실행하면, *json* 형태로 작성시간, 작성자, 트윗 내용 등을 가진 객체를 반환한다.

```
api = tweepy.API(auth)  
tweets = api.search(keyword) #keyword for searching
```

트윗 내용이 여러 줄이면 한 객체가 여러 개의 트윗으로 인식될 수 있기 때문에 'wn'을 공백으로 대체한다. 이로써 한 사람이 작성한 트윗이 한 개로 인식된다. *json* 객체 중 트

³ 1) <http://blog.naver.com/nonamed0000/220912854545>, 2) <http://rachelee.org/72>를 참고함.

⁴ <https://apps.twitter.com/>

윗 내용('text')만을 데이터 형태로 저장한다.

```
s = tweet.text.replace('\n', ' ')
data['text'] = s
```

트위터에서 크롤링한 우울감 관련 키워드는 'depressed' (1,500개), 'depression' (2,230개), 'anxiety' (2,250개) 등 총 **5,980**개 트윗을 모았고, 행복 관련 키워드는 'happy' (1,486개), 'comfortable' (2,141개), 'relaxed' (2,229개) 등 총 **5,856**개 트윗을 모았다.

3) 전처리

■ 중복 제거

트위터의 경우, RT(re-twit) 기능 때문에 트윗 중복 내용이 많아, 이를 제거하기 위해 엑셀의 데이터 기능 중 중복된 항목 제거를 이용한다.

■ 언어 통일성

SNS에서 영어로 검색하면, 영어는 공용어이기 때문에 아랍어, 러시아어, 일본어 등 다양한 결과가 나왔다. 이러한 언어들은 분석에 도움이 되지 않으므로 삭제한다.

■ 유니코드 인코딩

SNS 게시글은 Python 프로그램에서 UTF-8 인코딩으로 저장되지만, CSV 파일과 WEKA에서 실행시에 오류를 일으키기도 하므로 (이모티콘, Non-breaking space 등) 메모장으로 옮겨 ANSI 인코딩으로 변경하였다.

■ 광고

교묘하게 섞인 광고를 제거하여야 한다. SNS 상에는 "Please follow me", "If you're in trouble, please call ..."과 같은 팔로워 유입을 목표로 한 게시글이 상당히 많다.

6,680개의 우울감 관련 게시글을 전처리한 후 남은 데이터는 **825**개이고, 6,756개의 행복 관련 게시글을 전처리한 후 남은 데이터는 **813**개이다.

3. 결과

1) WEKA를 이용한 기계학습

■ 사용하기에 앞서

가중치를 동일하게 맞추기 위하여 depression 클래스의 게시글 **813**개, happy 클래스의 게시글 **813**개를 대상으로 실험하였다. CSV 파일을 사용하였기 때문에 텍스트를 인식하게 하기 위해서, 작은 따옴표(')는 밑줄(_)로, 큰 따옴표(")는 공백으로 변환한 뒤, *nominalToString* 필터를 사용해 Nominal 값으로 인식된 텍스트를 String으로 변환하였다. 그 후, *StringToWordVector* 필터를 이용해 다음과 같이 변환했다.

Viewer

Relation: AFFreal_comman - 복사본-weka.filters.unsupervised.attribute.NominalToString-Clast-weka.filters.unsupervised.attribute.StringToWordVector-R2-W1000-prune-ra

No.	1: class Nominal	2: # Numeric	3: #13reasonswhy Numeric	4: #2018 Numeric	5: #Depression Numeric	6: #Happiness Numeric	7: #WorldCup Numeric	8: #adhd Numeric	9: #aesthetic Numeric	10: #alone Numeric	11: #alternative Numeric	12: #amazing Numeric
1	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
9	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	depression	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	depression	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14	depression	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
15	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
17	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
18	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
19	depression	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20	depression	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0

■ Naïve Bayesian analysis

이후 NaiveBayes 분석 도구를 실행하면, 84.50%라는 정확도를 얻게 된다.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      1374           84.5018 %
Incorrectly Classified Instances    252           15.4982 %
Kappa statistic                    0.69
Mean absolute error                 0.1631
Root mean squared error             0.3519
Relative absolute error             32.6274 %
Root relative squared error         70.3884 %
Total Number of Instances          1626

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0.891    0.200    0.816     0.891    0.852     0.693    0.923    0.895    depression
          0.800    0.109    0.880     0.800    0.838     0.693    0.923    0.934    happy
Weighted Avg.   0.845    0.155    0.848     0.845    0.845     0.693    0.923    0.915

=== Confusion Matrix ===

  a  b  <-- classified as
724  89 |  a = depression
163 650 |  b = happy
|

```

■ Ranker method

Ranker method를 사용하면, 어떤 단어가 분석에 영향을 주었는지 알 수 있다. 상위에 포진한 단어들은 #depression, #anxiety, #iamhappy 등 크롤링 시작 단계에서 설정한 단어들이기 때문에, 이러한 직접적 표현이 없어도 우울감을 감지하는 프로그램을 만들려면 해당 단어들을 삭제하여야 한다.

=== Attribute selection 5 fold cross-validation (stratified), seed: 1 ===

average merit	average rank	attribute
0.183 +- 0.005	1 +- 0	66 #depression
0.089 +- 0.004	2 +- 0	436 anxiety
0.078 +- 0.004	3 +- 0	528 depression
0.067 +- 0.002	4.6 +- 0.8	127 #iamhappy
0.068 +- 0.002	4.8 +- 0.4	16 #anxiety
0.062 +- 0.003	6.4 +- 0.49	504 comfortable
0.058 +- 0.002	7.2 +- 0.75	63 #depressed
0.055 +- 0.008	7.8 +- 2.32	633 happy
0.048 +- 0.004	9.4 +- 1.36	227 #sad
0.046 +- 0.004	10 +- 1.26	527 depressed
0.047 +- 0.001	10.6 +- 1.02	113 #happy
0.044 +- 0.002	11.2 +- 0.75	112 #happiness
0.039 +- 0	13 +- 0	333 Happy
0.03 +- 0.003	14.4 +- 0.49	254 #suicide
0.03 +- 0.002	14.6 +- 0.49	174 #mentalhealth
0.023 +- 0.002	16.8 +- 0.75	136 #instagood
0.022 +- 0.002	18 +- 1.67	176 #mentalillness
0.021 +- 0.002	18.6 +- 1.96	98 #fun
0.021 +- 0.001	19.6 +- 1.5	242 #smile
0.021 +- 0.002	19.8 +- 3.06	25 #beautiful
0.02 +- 0.001	20.8 +- 1.17	38 #broken
0.019 +- 0.002	21.6 +- 2.42	319 Depression
0.017 +- 0.002	24.8 +- 2.48	9 #alone
0.017 +- 0.002	25 +- 4.56	823 relaxed
0.016 +- 0.002	26.6 +- 3.61	253 #suicidal
0.017 +- 0.002	26.8 +- 3.97	202 #photooftheday
0.016 +- 0.002	27.2 +- 3.43	138 #instalike
0.016 +- 0.001	28.4 +- 3.38	204 #picoftheday
0.015 +- 0.002	29.2 +- 3.6	476 birthday
0.014 +- 0.002	33 +- 6.42	14 #anorexia
0.014 +- 0.001	33.6 +- 5.08	267 #travel
0.014 +- 0.001	33.8 +- 2.86	170 #meme
0.014 +- 0.001	34.4 +- 4.5	67 #depressionquotes
0.013 +- 0.001	35.6 +- 4.13	163 #lonely

그리고 위 결과는 대문자, 소문자가 구분되어 있기 때문에 모두 소문자로 바꾸고, 해시태그 표시인 샵(#)이 있을 때와 없을 때도 구분하여 결과를 내보기로 하였다. 따라서 다음과 같은 추가 실험을 통해 결과가 개선될 수 있다: 1) Depression, happy, happiness 등의 직접적인 검색어 삭제, 2) 대문자를 소문자로 바꾸는 것에 따른 결과 변화, 3) 샵의 유무에 따른 결과 변화.

2) 결과 개선

① 직접적인 검색어 삭제

직접적으로 감정을 나타내는 해당되는 단어를 제거하였을 때, 기존의 84.50%에서 **65.07%**이라는 감소된 정확도를 나타냈다. 다시 Ranker method를 실행하면, sad, suicide, mentalhealth, fun, mentalillness, smile 등의 의미가 유사한 어휘들이 상위에 포진해 있다.

```
=== Attribute selection 5 fold cross-validation (stratified), seed: 1 ===
```

average merit	average rank	attribute
0.048 +- 0.004	1 +- 0	216 #sad
0.03 +- 0.003	2.4 +- 0.49	243 #suicide
0.03 +- 0.002	2.6 +- 0.49	163 #mentalhealth
0.023 +- 0.002	4.8 +- 0.75	125 #instagood
0.021 +- 0.002	6.2 +- 1.33	90 #fun
0.022 +- 0.002	6.2 +- 1.94	165 #mentalillness
0.021 +- 0.001	7 +- 1.1	231 #smile
0.021 +- 0.002	7.4 +- 2.42	22 #beautiful
0.02 +- 0.001	8 +- 1.41	35 #broken
0.017 +- 0.002	11 +- 1.67	7 #alone
0.016 +- 0.002	13.4 +- 3.5	242 #suicidal
0.017 +- 0.002	13.6 +- 3.38	191 #photooftheday
0.016 +- 0.002	13.8 +- 2.56	127 #instalike
0.016 +- 0.001	14.4 +- 3.01	193 #picoftheday
0.015 +- 0.002	15.2 +- 3.49	459 birthday
0.014 +- 0.002	19 +- 6.07	12 #anorexia
0.014 +- 0.001	19.2 +- 4.62	256 #travel
0.014 +- 0.001	19.6 +- 2.8	159 #meme
0.013 +- 0.001	20.6 +- 3.83	152 #lonely
0.013 +- 0.001	21.2 +- 4.58	249 #tbt
0.013 +- 0.002	22.8 +- 7.57	709 me
0.011 +- 0.001	27.8 +- 8.47	218 #sadness
0.011 +- 0.001	28.2 +- 5.46	223 #selfharm
0.011 +- 0.001	28.8 +- 4.07	11 #anime
0.011 +- 0.002	29.4 +- 9.22	244 #summer
0.012 +- 0.002	29.8 +-15.43	619 health
0.011 +- 0.002	30.2 +- 7.17	181 #pain
0.011 +- 0.003	31.8 +-12.42	87 #friends
0.011 +- 0.002	33 +-13.31	733 my
0.01 +- 0	34.8 +- 4.07	50 #cute
0.01 +- 0.001	36.6 +- 7.31	219 #sadquotes
0.01 +- 0.002	37.6 +-12.56	240 #style
0.01 +- 0.002	38 +-13.1	164 #mentalhealthawareness
0.01 +- 0.002	38 +-11.4	924 trying
0.01 +- 0.002	40.6 +-12.99	75 #fashion

② 대문자를 소문자로 바꾸는 것에 따른 결과 변화

```
lines = f.readlines()
for line in lines:
    line = line.lower()
    file.write(line)
```

②와 ③ 실험을 하기 위해서는 Python 프로그램을 이용해 파일을 전처리해주어야 한다. 전처리 후 기존 결과와 비교해보면, 데이터에 대문자가 포함되어 있을 때 84.50%였던 정확도가 **87.52%**가 되었다. 여기에 직접적인 검색어 제거(①)까지 수행하면, 65.07%에서

67.47%의 정확도로 향상되었다.

```

Correctly Classified Instances      1097          67.4662 %
Incorrectly Classified Instances    529          32.5338 %
Kappa statistic                    0.3493
Mean absolute error                 0.3384
Root mean squared error             0.4999
Relative absolute error             67.6713 %
Root relative squared error        99.9832 %
Total Number of Instances         1626

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.699	0.349	0.667	0.699	0.682	0.350	0.748	0.711	depression
	0.651	0.301	0.683	0.651	0.667	0.350	0.748	0.774	happy
Weighted Avg.	0.675	0.325	0.675	0.675	0.674	0.350	0.748	0.742	

③ 샵의 유무에 따른 결과 변화

해시태그를 일반 텍스트로 취급하였을 때, 정확도가 84.50%에서 **85.85%**로 약간 개선되었다. 여기에 직접적인 검색어 제거(①)까지 수행하면, 정확도가 65.07%에서 **67.06%**로 개선되었다.

```

=== Summary ===

Correctly Classified Instances      1085          67.0581 %
Incorrectly Classified Instances    533          32.9419 %
Kappa statistic                    0.3412
Mean absolute error                 0.3381
Root mean squared error             0.4989
Relative absolute error             67.616 %
Root relative squared error        99.7845 %
Total Number of Instances         1618

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.716	0.375	0.656	0.716	0.685	0.343	0.749	0.714	depression
	0.625	0.284	0.688	0.625	0.655	0.343	0.749	0.772	happy
Weighted Avg.	0.671	0.329	0.672	0.671	0.670	0.343	0.749	0.743	

②와 ③의 조건을 모두 적용하면, 정확도가 84.50%에서 **88.50%**로 크게 개선되었다. 여기서 직접적인 검색어 제거(①)를 적용하면, **69.10%**로 감소하지만, ①만 적용했을 때의 정확도인 65.07%보다 4% 가량 개선되었다. 추가적으로, Meme, 숫자, @태그, rt 등의 분석에 관계 없는 내용과 a, an, this 등의 관사를 제거하면, **68.29%**로 낮아진 것을 확인할 수 있다. 원인은 제대로 광고나 떠도는 글이 제거되지 않았거나, 위와 같은 토큰이 class 결정에 영향을 미친 탓일 것이다.

3) 결과 해석

이 실험의 흥미로운 점은 Rank method를 활용해 우울감과 관련된 현상이나 언어를 알아낼 수 있다는 것이다. 다음의 세 가지 초점으로 결과를 해석해볼 수 있다.

■ 감정과 관련된 단어

당연하게도, 슬픔 (sad, sadness), 외로움 (alone, lonely), 증오 (hate) 등의 주변 어휘들이 상위에 나타났다. 이러한 어휘들은 우울감을 호소하는 환자들의 진단에 활용될 수 있는 언어 표현이라고 볼 수 있다. 과거의 나를 의미하는 tdt (Throwback Thursday)도 상위에 랭크 되었는데, 과거에 대한 향수를 불러오는 사진을 게시하며 자주 사용하는 단어라고 한다. 행복과 우울 모두와 관계될 수 있는 단어로, 상위에 있는 점이 특이하다.

■ 생일

검색어 'happy'의 경우 'happy anniversary', 'happy birthday'와 같이 자주 사용되기 때문에 'birthday'가 상위에 랭크 되었다. 이 문제점은 크롤링 과정의 한계점(특정 검색어로 크롤링해야 하는 문제)으로 추후 개선되면 기계학습의 성능을 향상시킬 수 있을 것이다.

■ 동반되는 증상

자살기도 (suicide, suicidal), 자해 (selfharm), 고통 (pain), 스트레스 (stress) 등의 토큰들은 우울감에 선행하거나, 혹은 뒤따르는 중요한 증상들을 나타낸다. 특히, 식욕부진 (anorexia)은 우울증이나 암, 결핵 등의 질환에 동반하는 것으로 알려져 있다 (서울대학교병원). 약 1600개 데이터 중 31회 출현하여 많은 사람들이 우울감과 식욕 부진으로 이중 고통 받고 있다는 사실을 알 수 있다.

Ranked attributes:

0.05132	714 sad	0.0131	734 selfharm
0.03486	816 suicide	0.01305	800 stress
0.02954	540 mentalhealth	0.01258	404 i
0.02781	86 birthday	0.01247	29 anime
0.02577	67 beautiful	0.01247	541 mentalhealthawareness
0.02499	430 instagood	0.01186	539 mental
0.02202	770 smile	0.01184	895 trying
0.02134	542 mentalillness	0.01164	360 hate
0.02016	19 alone	0.01123	631 pain
0.01801	107 broken	0.01075	428 instadaily
0.01801	653 picoftheday	0.01048	716 sadness
0.01784	650 photooftheday	0.00996	231 emo
0.01753	432 instalike	0.00996	978 worthless
0.01728	815 suicidal	0.00996	717 sadquotes
0.01683	309 fun	0.00992	11 aesthetic
0.01563	31 anorexia	0.0097	264 fashion
0.01436	502 lonely	0.0095	896 tumblr
0.01398	817 summer	0.00936	182 death
0.01373	886 travel	0.00936	809 style
0.0131	835 tbt	0.00933	715 sadedits

4. 결론

우울감으로 고통 받는 주변 사람들을 보고서 시작하게 된 프로젝트였지만, 데이터를 수집하는 과정에서 수도 없이 많은 사람들이 자살 기도, 우울증, 자해 등으로 괴로워하며 친구나 의사를 통해 도움을 받는 것이 아닌, SNS 상에서 조언을 구하고, 자해 사진, 우울한 글('sad quotes', 'depression quotes' 역시 상위 랭크를 차지했다)을 게시하는 쪽을 선택하여 자신의 우울감을 드러내는 것을 보게 되었다. 실제로 의사를 찾아가거나 친구에게 털어놓는 것이 사회적으로 부담된다는 의미일 것이다. 어렵게 자신의 심리 상태를 나에게 털어놓은 친구들을 떠올리며, 추후 이 프로젝트가 발전되어서 온라인으로 심리상담을 받거나, AI를 활용해 우울감을 조기 진단할 수 있는 기술이 만들어지기를 진심으로 기원해본다.

참고 문헌

- 김경욱. (2018년 06월 02일). 샤이니 종현, 생전 마지막 게시물 "네가 아파하지 않길 기도해". "화이트페이퍼".
- 김기성. (2007년 04월 18일). 美총기사건 범인 교포학생 `경악`..증오범죄 우려(재종합). "이데일리".
- 김기홍한경식. (2018). 기계학습 및 소셜미디어 이미지 기반 우울증 분석 모델 개발. "한국통신학회", 페이지: 832.
- 대중문화타임스포츠한국. (2017년 12월 20일). 정신 전문가 "샤이니 종현 팬들 베르테르 효과 대처하려면". "스포츠한국".
- 서울대학교병원. (날짜 정보 없음). 식욕 부진. "서울대학교병원 의학정보".
- 선미경. (2017년 12월 19일). 故종현, 부검 않고 21일 발인..상주는 샤이니 멤버들[종합]. "OSEN".
- 연합뉴스. (2007년 04월 18일). <美 총기난사사건은 '외톨이'의 억눌린 분노 때문>(종합). "연합뉴스".
- 최홍구황인준. (2012년 11월). 트위터 문서 분석을 통한 감정 기반의 음악 추천 시스템. "정보과학회논문지", 페이지: 762.