

# Deep Learning Side Project

## mod group one

- Yoon Sangseok
- Kim Hyunji
- Jeong Bongki

## 국가 청원 글 학습 모델

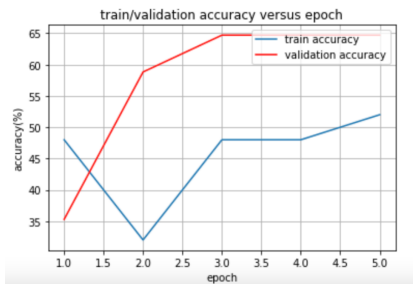
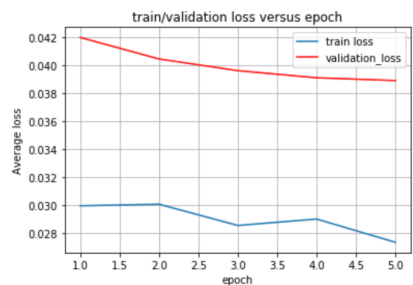
크롤링으로 데이터 수집을 하고 그 데이터를 전처리 작업을 거치 후에 cleaned data를 기반으로 카테고리 분류를 하고 기존에 학습한 데이터들을 기반으로 새로운 청원글이 올라왔을 때 해당 글이 청원 목표수를 넘길 수 있는지 예측하는 모델



BeautifulSoup

pandas

	category	content	count	end	start	title
0	행정	안녕하십니까 대통령님.Vn저는 경기도 양주에 살고 있으면 경기도 양주에 소규모 ...	686	2021-01-09	2020-12-10	대통령님 코로나 집합금지로 인해 소상공인 자영업자 분들이 신용불량 위기에 처해 있습니다.
1	기타	안녕하세요.Vn저는 12월3일 mbc에서 방송된 '롯데호텔성의 내부고발자에대한 보복...	6,024	2021-01-09	2020-12-10	MBC에서도 보도된 롯데호텔성의 내부고발자 죽이기를 막아주세요
2	안전/환경	저는 코로나 전담병원 *****에서 코로나 환자를 보고있는 간호사이자 두 아이의 엄마...	1,619	2021-01-09	2020-12-10	아이들을 지켜주세요
3	반려동물	오늘 새벽(6일)에서 7일 날아가는 12시 경 지희 강아지가 세상을 떠났습니다.V...	868	2021-01-09	2020-12-10	동물법에 의뢰하고 포항 및 사람과 동등한 처벌을 받을수있도록 해주세요
4	육아/교육	평범한 시민으로서 보다 못해 청원을 남깁니다.매년 반복되는 돌봄 사태가 무슨 일인가...	5,920	2021-01-09	2020-12-10	현실감각, 현장감각 전혀 없는 비전문가 유은혜 교육부 장관은 사퇴하라.
5	기타	이제는 마지막입니다... VnVn먼저 정부의 코로나19 극복을 위한 노력에 감사...	506	2021-01-09	2020-12-10	실내체육시설 이제는 정말 마지막입니다...
6	정치개혁	역사상 가장 큰 동력으로 검찰개혁을 진행하고 있는 현재, 해당 검찰조직은 정치권 및 ...	1,478	2021-01-09	2020-12-10	불가피한 검찰개혁
7	안전/환경	양평군민과 수도권시민 다 죽겠다 "용문산 사격장 폐쇄하라"VnVn양평군민과 수도권...	2,174	2021-01-09	2020-12-10	양평군민과 수도권시민 다 죽겠다 "용문산 사격장 폐쇄하라"
8	반려동물	12월6일 오후4시 13살된 지희 강아지 동이가. 요로결석으로 인해 급수술을 받았...	1,222	2021-01-09	2020-12-10	반려동물의료사고
9	외교/통일/국방	<이 시국에 마스크도 착용하지 않고 덴스파리를 발인 주한미군에 대한 엄중한 처벌을 요...	1,083	2021-01-09	2020-12-10	이 시국에 마스크도 착용하지 않고 덴스파리를 발인 주한미군에 대한 엄중한 처벌을 요...
10	안전/환경	12/8일 오늘부터 코로나 단계가 2.5단계로 극상한 가운데 백화점 화장...	280	2021-	2020-	코로나2.5단계 백화점 직원들무 뜻아갑니다



```
textCNN(  
    (embed): Embedding(165, 100)  
    (convs): ModuleList(  
      (0): Conv2d(1, 10, kernel_size=(3, 100), stride=(1, 1))  
      (1): Conv2d(1, 10, kernel_size=(4, 100), stride=(1, 1))  
      (2): Conv2d(1, 10, kernel_size=(5, 100), stride=(1, 1))  
    )  
    (dropout): Dropout(p=0.6, inplace=False)  
    (fc): Linear(in_features=30, out_features=2, bias=True)  
)  
Train Epoch: 1    Loss: 0.02997595548629761    Accuracy: 48.0%  
Valid Epoch: 1    Loss: 0.041975119534660786    Accuracy: 35.29411697387695%  
model saves at 35.29411697387695 accuracy  
Train Epoch: 2    Loss: 0.030087029933929442    Accuracy: 32.0%  
Valid Epoch: 2    Loss: 0.040440142154693604    Accuracy: 58.82352828979492%  
model saves at 58.82352828979492 accuracy  
Train Epoch: 3    Loss: 0.02857349157333374    Accuracy: 48.0%  
Valid Epoch: 3    Loss: 0.03960788951200597    Accuracy: 64.70587921142578%  
model saves at 64.70587921142578 accuracy  
Train Epoch: 4    Loss: 0.029030663967132567    Accuracy: 48.0%  
Valid Epoch: 4    Loss: 0.03909907270880306    Accuracy: 64.70587921142578%  
Train Epoch: 5    Loss: 0.027375001907348633    Accuracy: 52.0%  
Valid Epoch: 5    Loss: 0.038902713971979475    Accuracy: 64.70587921142578%  
-----  
Test Epoch: 5    Loss: 0.039124995470047    Accuracy: 50.0%
```



matplotlib

PyTorch

	token_final	label
0	[현실, 코로나, 국민, 집합, 금지, 대통령, 님, 코로나, 집합, 금지...	No
1	[동생, 회사, 합의, 협박, 증거, 대기업, 애서도, 보도된, 롯데, 황성, 의,...	Yes
2	[코로나, 지희, 병원, 환자, 병실, 아이, 의료, 아이, 들, 을, 지켜, 주,...	Yes
3	[지희, 동물병원, 아이, 수술, 시간, 아침, 동물법, 예, 의뢰사, 고, 포함,...	No
4	[따입, 시간, 교육부, 초등, 장관, 현실, 감각, 현장, 감각, 전혀, 없는, ...	Yes
5	[입증, 마스크, 운영, 시설, 이제, 실내, 체육, 실내, 체육, 시설, 이제, ...	No
6	[검찰, 조직, 헌법, 현재, 정치, 불가피한, 검찰, 개혁]	Yes
7	[군민, 사격장, 양평, 용문산, 지난, 수도권, 양평, 군민, 과, 수도권, 시민...	Yes
8	[병원, 아이, 지희, 수술, 의사, 내일, 반려동물, 의뢰사, 고]	Yes
9	[주한, 미군, 방역, 국민, 코로나, 이, 시국, 예, 마스크, 도, 착용, 하지...	Yes
10	[백화점, 직원, 경우, 실사, 재고, 조사, 코로나, 2, ,, 5, 단계, 백화...	No
11	[평가, 훈련, 인원, 청원, 특전사, 통제, 특수전사령부, 7, 공수, 여단, 훈...	No
12	[교습, 지도, 점검, 과외, 수업, 아파트, 과외, 방, 은, 코로나, 로부터, ...	No
13	[이전, 거한, 조지, 자과, 전복, 거참출자, 이라는, 면친, 오, 거참출자, 오...	Yes

# 기술 스택

## plowchart

- BeautifulSoup4
  - 데이터 수집 (크롤링)
- Pandas
  - 데이터 분석
- KoNLPy
  - 자연어 처리 (한국어 정보처리)
- Pytorch
  - 딥러닝
- Matplotlib
  - 데이터 시각화

