2023 Data competition

PRESENTATION

2023 데이터과학부 학술제

가짜 뉴스 탐지기

fake news reader

3학년 10조/ 진용훈 이우리 박다윤

CONTENTS

중간 발표 주제 - 뉴스 기사 레이블 분류

선정 이유

- 데이콘 해커톤을 통해 NLP 분류 task 첫 연습
- Sentence Bert 모델을 활용하여 뉴스 기사의 레이블 분류에서 fine-tuninng 시도

category	info
0	Business
1	Entertainment
2	Politics
3	Sports
4	Tech
5	World

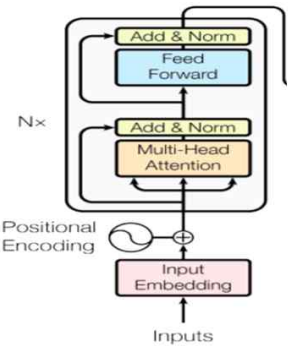
중간 발표 주제 - 뉴스 기사 레이블 분류

But,

뉴스 분류를 응용하는 것에 한계를 깨달음

=> bert 계열의 모델에 대한 지식을 활용하여 실용적인 과제를 찾고자 함

=> '가짜 뉴스 탐지' 를 목표



2023 Data competition

주제 선정 이유

가짜 뉴스의 문제점 ⚠

사실왜곡, 사회 혼란 -> 무고한 피해자 발생

[디지털 이코노미] 가짜뉴스가 주가에 미치는 영향 진짜뉴스의 3배

생글생글 756호

(55) 디지털경제와 가짜뉴스

공정사회

기획 연재

가짜뉴스가 금융·정치·경제 디스

유의미한 정책설계를 위해서는 소

백악관에 폭발 사고가 발생해 오바마

4000번 넘게 리트윗됐고, 수십만 명

춤하던 시장 반응은 곧 궤도를 이탈하

를 제안하는 데이터 회사들이 만든 지

URL

"가짜뉴스 무차별 생산, 유포에 피해사례 속출…피해자 고통만 가중"

[기획연재-공정사회의 적, 가짜뉴스]⑤'나는 가짜뉴스 피해자다'

기사입력 2023.09.11 06:00 | 한국NGO신문 기획취재팀

가 가

대한민국에 '가짜뉴스'가 범람하고 있다. 특정 분야만 아니라 정치, 경제, 사회, 문화, 방송·연예 등까지 광범위하게 '가짜뉴스'가 퍼지면서 사실을 왜곡, 호도하는 것은 물론 사회적 혼란과 부역을 야기하고 있다. 특히

PHOTO 뉴스

주제 선정 이유

AI 가짜사진에 美 월스트리트도 발칵..!

▲ 허위로 밝혀진 '펜타곤 화재'

지난 5월 22일, 미국 펜타곤 청사 근처에서 대형 폭발이 발생했다고 주장하는 이미지가 큐어넌, 페이스북에 공식 계정 마크를 달고 있는 수많은 계정 등 sns를 통해 전세계로 확산되었다. 소식이 전해지자 **美 S&P 500 지수가 한때 0.3%, 다우 존스 30(DJI) 산업평균지수가 약 80p 하락하는 등 증시에 영향이 갔고, 美 국채와 금값이 잠시 상승하는 현상이 나타났다.**

AI가 정치판도 뒤흔든다...

▲ 도널드 트럼프 전 대통령이 자신의 소셜미디어에 올린 딥페이크 동영상

생성형 AI는 선거판의 새로운 복병으로도 떠오를 것으로 보인다. 기존 공식 선거 운동은 실제 영상만을 사용했다면, 미래에는 AI를 활용한 다양한 전략이 나타날 것이다. 이는 현재 정치 분야에 강한 AI 기술 이용 규칙이 없기 때문이기도 하다. 내년 11월 미국 대선이 AI 가짜 영상에 의한 흑색선전에 오염될 것이란 우려가 고조되고 있다. **AP통신은 14일 "AI가 2024년 대선에 위험을 초래할 수 있다"라고 전했다.**

2023 Data competition

목표:가짜 뉴스 탐지기 제작

가짜뉴스 피해 해결 방안

목표:
진짜 뉴스와 가짜 뉴스가 무엇인지 판별하는 분류 모델을 만들어보자

예상 효과:
-사용자들이 신뢰할 수 있는 정보에 빠르게 액세스 할 수 있도록 도와주는 조수 역할

-정보 신뢰성과 사회적 안정성을 유지하는 데 중요한 도구로 활용 가능

데이터 정의

kaggle의 Fake and real news dataset

title, text, subject,date를 포함한 약 2만행의 true.csv와 fake.csv

	title	text	subject	date
2	As U.S. budget figh	WASHINGTON (Reute	politicsNews	31-Dec-17
3	U.S. military to acce	WASHINGTON (Reute	politicsNews	29-Dec-17
4	Senior U.S. Republi	WASHINGTON (Reute	politicsNews	31-Dec-17
5	FBI Russia probe he	WASHINGTON (Reute	politicsNews	30-Dec-17
5	Trump Is So O	On Christmas day	News	29-Dec-17
6	Pope Francis Ju	Pope Francis usec	News	25-Dec-17
7	Racist Alabama	The number of ca	News	25-Dec-17
8	Fresh Off The (	Donald Trump sp	News	23-Dec-17
9	Trump Said So	In the wake of ye	News	23-Dec-17
10	Former CIA Dir	Many people hav	News	22-Dec-17

LIAR Dataset

truthfulness, subject, context/venue, speaker, state, party, and prior history 으로 구성된 12,836 행의 .tsv파일

	2635.json	false	Says the Annies List political group supports third-trimester abortions on demand.	abortion	dwayne-bohac
0	10540.json	half-true	When did the decline of coal start? It started...	energy,history,job-accomplishments	scott-surovell
1	324.json	mostly-true	Hillary Clinton agrees with John McCain "by vo...	foreign-policy	barack-obama
2	1123.json	false	Health care reform legislation is likely to ma...	health-care	blog-posting
3	9028.json	half-true	The economic turnaround started at the end of ...	economy,jobs	charlie-crist
4	12465.json	true	The Chicago Bears have had more starting quart...	education	robin-vos

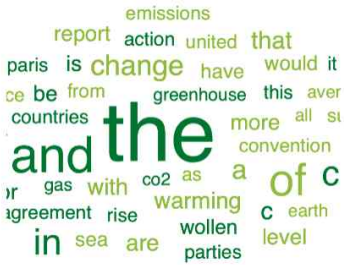
COVID19 Fake News Detection in English

데이터 소스는 트위터, 페이스북, 인스타그램 등 다양한 소셜 미디어 플랫폼으로 구성된 8560 행의 데이터셋

	id	tweet	label
0	1	The CDC currently reports 99031 deaths. In gen...	real
1	2	States reported 1121 deaths a small rise from ...	real
2	3	Politically Correct Woman (Almost) Uses Pandem...	fake
3	4	#IndiaFightsCorona: We have 1524 #COVID testin...	real
4	5	Populous states can generate large case counts...	real

데이터 전처리

LEMMATIZATION 및 클렌징



STOP WORDS 제거

텍스트에 자주 쓰이지만 중요하
지 않은 단어들 제거
EX) (A, THE,I ..)

소문자 변환

자연어 처리 작업에서 노이즈로
작용할 수 있는
대문자들 소문자로 변환

제목과 텍스트 합치기

모델의 INPUT으로 한 번에
들어갈 수 있게 제목과
텍스트를 합침

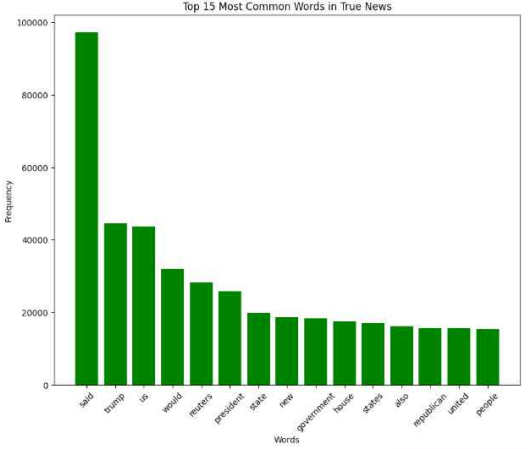
텍스트 길이 조절

분류에 악영향이 가는 너무 짧은
텍스트와
토큰 길이 제한을 유발하는 너무
긴 텍스트 제거

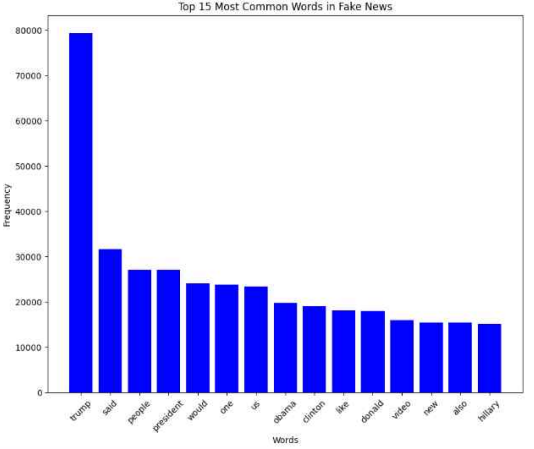
데이터 시각화

가정 : 진짜 뉴스의 단어와 가짜 뉴스의 단어의 차이가 있을거라고 생각

진짜 뉴스



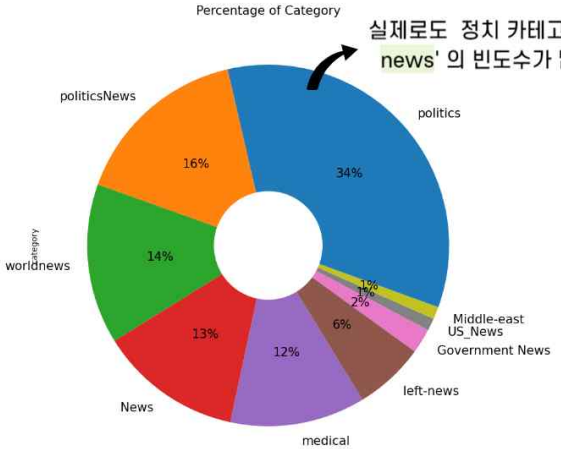
가짜 뉴스



But, 데이터셋에 대량의 정치 데이터셋이 존재

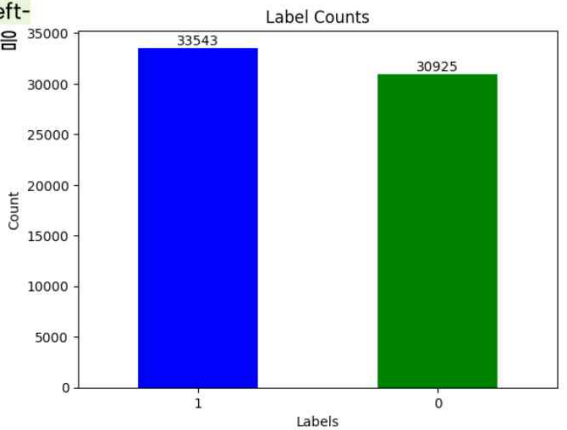
=> 정치관련 이름 및 단어들의 빈도수 多 , 비교 결과 유의미한 차이가 있다고 보기 어려웠음

데이터 시각화



카테고리별 빈도수

politics, medical, worldnews 등
다양한 카테고리 존재



각 라벨의 개수

0: True => 30925

1: Fake => 33543

모델 선정 과정

SBERT(중간평가 때 사용한 모델)

SBERT는 BERT를 샴 네트워크,
트리플렛 네트워크 구조로 파인튜닝한 모델

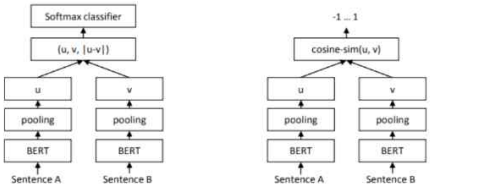


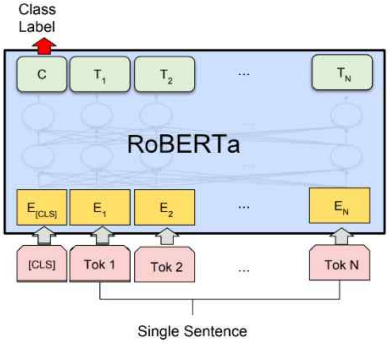
Figure 1: SBERT architecture with classification objective function, e.g., for fine-tuning on SNLI dataset. The two BERT networks have tied weights (siamese network structure).

Figure 2: SBERT architecture at inference, for example, to compute similarity scores. This architecture is also used with the regression objective function.

교체 이유: SBERT는 문장 쌍의 유사도를 구하는데 적합한 모델인데 주제의 변경으로 인해 다른 모델이 필요해짐

최종선정: RoBERTa

선정이유: 해커톤 때 분류로 해결한 사람들 중 우리가 사용한 모델보다 성능이 좋았던 모델



모델간 성능 비교 (ACC)

RNN

```
Epoch 8/10
1411/1411 [==============================] - 11s 8ms/step - loss: 0.0172 - mcc: 0.9
890 - val_loss: 0.2456 - val_mcc: 0.8994
Epoch 9/10
1411/1411 [==============================] - 11s 8ms/step - loss: 0.0177 - mcc: 0.9
873 - val_loss: 0.2565 - val_mcc: 0.8908
Epoch 10/10
1411/1411 [==============================] - 11s 8ms/step - loss: 0.0932 - mcc: 0.9
237 - val_loss: 0.3593 - val_mcc: 0.7078
Model: "sequential_4"
```

Layer (type)	Output Shape	Param #
embedding_4 (Embedding)	(None, 100, 30)	900000
simple_rnn_4 (SimpleRNN)	(None, 100)	13100
dense_4 (Dense)	(None, 1)	101

```
Total params: 913201 (3.48 MB)
Trainable params: 913201 (3.48 MB)
Non-trainable params: 0 (0.00 Byte)
```

0.9237

LSTM

```
Epoch 8/10
1411/1411 [==============================] - 27s 19ms/step - loss: 0.0108 - accurac
y: 0.9962 - val_loss: 0.2398 - val_accuracy: 0.9599
Epoch 9/10
1411/1411 [==============================] - 24s 17ms/step - loss: 0.0097 - accurac
y: 0.9965 - val_loss: 0.2068 - val_accuracy: 0.9643
Epoch 10/10
1411/1411 [==============================] - 24s 17ms/step - loss: 0.0089 - accurac
y: 0.9970 - val_loss: 0.2320 - val_accuracy: 0.9611
Model: "sequential_8"
```

Layer (type)	Output Shape	Param #
embedding_8 (Embedding)	(None, 100, 30)	900000
dropout_6 (Dropout)	(None, 100, 30)	0
lstm_1 (LSTM)	(None, 100)	52400
dropout_7 (Dropout)	(None, 100)	0
dense_10 (Dense)	(None, 64)	6464
dropout_8 (Dropout)	(None, 64)	0
dense_11 (Dense)	(None, 1)	65

```
Total params: 958929 (3.66 MB)
Trainable params: 958929 (3.66 MB)
Non-trainable params: 0 (0.00 Byte)
```

0.9611

GRU

```
Epoch 8/10
1411/1411 [==============================] - 23s 16ms/step - loss: 0.0099 - accurac
y: 0.9969 - val_loss: 0.2315 - val_accuracy: 0.9652
Epoch 9/10
1411/1411 [==============================] - 23s 16ms/step - loss: 0.0080 - accurac
y: 0.9972 - val_loss: 0.2599 - val_accuracy: 0.9644
Epoch 10/10
1411/1411 [==============================] - 23s 17ms/step - loss: 0.0069 - accurac
y: 0.9979 - val_loss: 0.2718 - val_accuracy: 0.9635
Model: "sequential_9"
```

Layer (type)	Output Shape	Param #
embedding_9 (Embedding)	(None, None, 64)	1920000
dropout_9 (Dropout)	(None, None, 64)	0
gru_1 (GRU)	(None, 64)	24960
dropout_10 (Dropout)	(None, 64)	0
dense_12 (Dense)	(None, 16)	1040
dropout_11 (Dropout)	(None, 16)	0
dense_13 (Dense)	(None, 1)	17

```
Total params: 1946017 (7.42 MB)
Trainable params: 1946017 (7.42 MB)
Non-trainable params: 0 (0.00 Byte)
```

0.9635

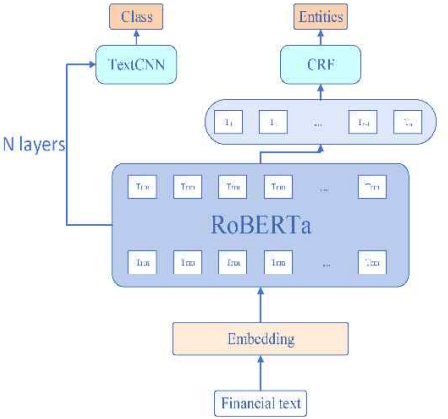
최종선정 모델: RoBERTa

RoBERTa란?

: Robustly optimized BERT approach,
기존의 BERT가 가진 몇가지 한계점(EX. underfitting)들을 개선한 모델

BERT에서 수정된점:

1. 더 많은 데이터를 사용하여 더 큰 배치로 학습
2. NSP 제거
3. 더 긴 시퀀스로 학습
4. dynamic masking



평가지표 최종 선정 -MCC

MCC란?

"Matthews Correlation Coefficient"의 약어로,
분류 모델의 성능을 측정하기 위한 통계적 지표

MCC의 적합성

1. AUC는 클래스 불균형이 심하면 높은 특이도를 갖고 이로 인해 모델 성능 과대평가 우려.
   반면 MCC는 클래스의 불균형에 민감하지 않아 적합

```
Matthews correlation coefficient (MCC) : 0.9820008750423032
True positive : 6631
True negative : 6147
False positive : 38
False negative : 78
Eval Loss : 0.040534075268160004
```

2. 4개의 혼동 행렬 범주의 균형 비율을 고려 -)
   이진 분류를 평가할 때 더 많은 정보 제공

3. 데이터의 불규칙성을 반영.
   이진 분류에서는 클래스 사이의 불규칙성 고려가 중요
   MCC는 실제 클래스 간의 관계를 고려하므로 모델의 성능을 정확히 평가 가능

최종 결과

테스트 결과

입력 텍스트: donald trump president scary prospect sane americans politicians gop democrat alike way behind bloviating fool crazy dangerous violent supporters california governor jerry brown.....

오약: 트럼프가 대통령이 되어야한다는 제목과
전혀 상관없는 내용을 가진 가짜뉴스

```
test_data = test['clean_message']
predictions, raw_outputs = model.predict(test_data[12891])
```

```
if predictions[0]==0:
  print("News is Real")
else:
  print("News is False")
```

```
News is False
```

가상 분류기 구축

뉴스 텍스트 입력

founder and co-editor of
21stCenturyWire.com.READ MORE SYRIA
NEWS AT: 21st Century Wire Syria Files

가짜 뉴스 확인하기

가짜 뉴스로 예측됩니다.

FINAL

가짜뉴스탐지기 활용 방안

1.뉴스리터러시 강화 교육
학생들이 비정상적인 정보를 구분하고,
미디어 소비에 있어서 비판적 사고 능력을
강화하는 데 활용

2.투자 결정 보조
정보의 신뢰성을 검증하고 투자 결정을 내릴 때 불필요
한 위험을 피할 수 있음

출처

모델 관련 논문

https://www.analyticsvidhya.com/blog/2022/03/a-brief-overview-of-recurrent-neural-networks-rnn/

http://colah.github.io/posts/2015-08-Understanding-LSTMs/

https://arxiv.org/abs/1907.11692

관련 뉴스 기사

https://sgsg.hankyung.com/article/2022052760561

http://www.ngonews.kr/news/articleView.html?idxno=144060