

# High Quality NEWS classification Service

With Deep-Learning

Date : 2020-09-04

MIAMI

김지현 김수연 김현우 이다정 심우열

## 일정 계획

[illegible]

# CONTENTS



## 1.개요

- 1) 배경
- 2) 분석 프로세스



## 2.데이터 수집 및 탐색

- 1) 데이터 수집
- 2) 데이터 탐색 및 전 처리



## 3.데이터 레이블링

- 1) 기존 연구 조사
- 2) 레이블링



## 4.모델 개발

- 1) 기존 연구 조사
- 2) 사용 모델
  - 2-1) LSTM
  - 2-2) Kogpt2
  - 2-3) KorBERT
- 3) 실험 결과



## 5.기대효과 및 시각화

- 1) 기대효과
- 2) 웹 서비스



# 개요



# 개요

## 배경

- 조회수를 늘리기 위한 은어, 비속어, 맥락 없는 내용 등 저품질 기사의 팽배 현상. 사실을 전달해야하는 기사의 본질을 흔들며 기사에 대한 신뢰도를 떨어뜨림

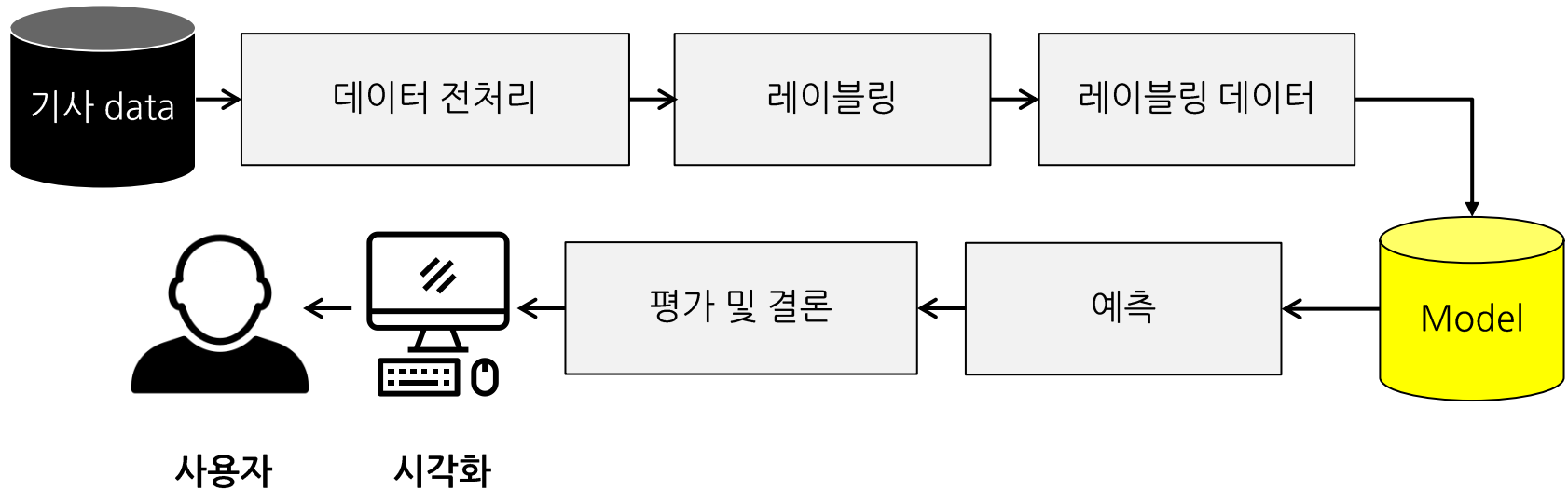


기사 품질을 측정하는 모델을 개발하여 저품질 기사 양산을 억제하고자 함

# 개요

## 분석 프로세스

- 아래의 분석 프로세스를 이용하여 프로젝트 진행



✓ 기사 data를 목적에 맞게 가공한 후 모델 설계, 이후 예측과 평가를 통해 최종모델을 웹에 구현하여 사용자에게 서비스 제공



# 데이터 수집 및 탐색



# 데이터 수집 및 탐색

## 데이터 탐색

- 매체 별 기사 서술방식이 상이해 공통된 특징을 찾아내고 서로 다른 부분을 파악하기 위함

컬럼명	예시	비고
기사번호	299918965 ~ 350853441	
매체 명	강원일보, 경기일보 등 (총 19개)	
매체유형	온라인 / 오프라인	온라인만 존재
매체구분	지역지 / 중앙지	지역지만 존재
발행일시	2019-03-21 12:12:12	
기사 카테고리	종합 / 경제 / 사회 / 스포츠 / 정치 / 교육	종합으로만 구성된 매체 존재
기사 대중소	소(0-999) / 중(1000-1499) / 대(1500-)	
제목	105만 고양시 발전 이끌어온 서기관 6명, 아름다운 퇴장	
내용	송주현 기자 왼쪽 위부터)김운용, 이현옥, 이흥민, 노양호, 정종현, 신승일 WrWn특례시 지정을 앞둔 105만 고양시의 발전을 위해 땀과 열정을 쏟아온 6명의 공직자(서기관)들이...	



# 데이터 수집 및 탐색

## 데이터 전처리

- ▶ 개별특성과 공통특성을 고려해 데이터 전처리 시행

▲최평균 S&T그룹 회장 “자기희생으로 S&T그룹 도약 이끌

자”WrWnWrWn최평균 (S&T그룹 회장) WrWnWrWn최평균 S&T그룹 회장은 3일 전 임직원들에게 “올해는 성장을 위한 구상보다 유비무환의 겸허함이 더욱 절실한 상황이다”며 ‘아무도 하지 않으면 내가 한다’는 자기희생을 주문했다.WrWnWrWn최 회장은 이날 ‘2019년 신년사’를 통해 “올해 기업경영이 치열한 글로벌 경쟁구도, 열강들의 경제전쟁, 국내 정치경제 환경들로 인해 바람 앞의 등불처럼 위태롭다”면서 이같이 강조했다. 특히 “자기희생은 이 시대가 요구하는 진정한 혁신의 모태이자, S&T 100년 역사를 이끄는 저력이 될 것으로 확신한다”면서 위기 극복을 위해 희생의 중요성을 강조했다. WrWnWrWn나아가 “최악의 상황을 고려해 단기간은 비관적으로 현상을 바라보면 문제가 보이고, 이를 세대가 공감하는 소통방식으로 용기 있게 부딪치면서 해답을 찾아 ‘생각 즉시 행동’하라는 것이 근원적 변화의 핵심”이라면서 변화와 혁신의 화두에 대해 구체적인 방안을 제시하기도 했다.WrWnWrWn이명용 기자 WrWn

경남신문<“대우조선해양·삼성중공업·S&T그룹 신년사 들어보

니”,2019.01.13.,정기홍기자> 중 일부

기사와 연관 없는  
특수문자 제거

괄호의 존재가 안에 내용이  
의미가 없다고 기계가  
판단하고자 함

단순한 공백을 구분하는  
개행문자 ‘WrWn’ 제거

“”는 인용문에 자주 사용되며  
제거시 의미해석에 방해를 줄  
수 있음

○ :보존 □ :제거



# 데이터 레이블링

---

# 데이터 레이블링

## 기존연구조사

- 논문을 토대로 고급기사의 특성을 파악하여 기사들을 직접 Labeling하고자 함

〈 고급 기사를 분류하기 위해 참고한 논문 〉

논문	저자	발행연도	내용
지상파 3사의 미디어 신뢰도와 TV뉴스 이용에 관한 연구	성동규, 박상호	2005	뉴스채널 선호집단에 따른 미디어 신뢰도가 미디어 이용에 미치는 영향차이를 '공정성', '정확성', '정보풍부성'의 판단 근거로 제시
뉴스평가지수의 개발과 적용	박재영, 이완수	2010	갈등사안여부, 관점제시여부, 이해당사자수, 투명취재원 수 등을 토대로 고급뉴스의 기준을 선정, 이진 분류 방법론 제시
방송뉴스의 언어와 표현 - 뉴스언어의 객관성과 공정성을 중심으로	박재영, 이완수	2013	표준어나 맞춤법에 맞지 않는 표현은 대중의 언어사용 관행을 잘못된 방향으로 이끔. 또한 기사는 사실정보에 기초하여야 하며 기사의 주관을 배제하여야 함
방송 법조뉴스의 품질 연구 정확성, 심층성, 불편부당성을 중심으로	박성호, 윤형민	2016	취재원이 실명으로 제시되면 정보의 출처를 검증할 수 있어 기사의 정확성이 높으며, 익명으로 제시될 경우 기사에 담긴 정보가 얼마나 근거가 있는지 알 수 없으므로, 그 수가 많을수록 기사의 정확성은 의심스러움. 사실이나 근거가 확실히 확인되지 않은 경우와 술어의 주체가 누구인지 확실히 표시되지 않은 경우 정확성이 낮음
방송뉴스의 단독 보도 품질	유수정, 오건호	2020	객관성은 투명 취재원의 활용으로, 다양성은 기사 내 등장한 이해당사자 수와 관점 유형으로, 전문성은 기사의 분석력과 기사 유형으로 측정. 투명 취재원 적시는 정확성을 높이는 요소로 평가됨

- ✓ 기사 전문에 나타나는 서술어, 명사 등 모든 표현을 확인하고 분류를 진행
- ✓ 최대한 많은 Data를 Labeling 하고자 했으며 Labeling Day를 설정해 총 1,003개의 뉴스 기사를 분류하여 Data Set 구축

# 데이터 레이블링

## 레이블링

### ➤ 레이블링 데이터 주요 변수 소개



#### 투명 취재원

이름이 명확히 밝혀진 경우



#### 익명 취재원

이름이 명확히 밝혀지지 않은 경우



#### 서술어 수

총 서술어 수



#### 주관적 서술어 수

주관이 개입된 서술어 수



#### 미확인 서술어 수

추측성 서술어 수



#### 비문

비표준어, 오타

# 데이터 레이블링

## 레이블링

### ➤ 기사를 품질기준\*에 따라 레이블링 데이터 생성

구예리 기자 유류세 인하 효과 등으로 경기지역 소비자물가 상승률이 5개월 만에 가장 낮은 수준을 기록했다. 경인지방통계청이 31일 발표한 '12월 소비자물가 동향'을 보면 이달 소비자물가지수는 1년 전보다 1.3% 상승했다. 지난 7월 1.2%를 기록한 뒤 5달 만에 가장 낮은 상승률이다. 품목 성질별로 보면 농·축·수산물 5.5% 올라 전체 물가를 0.39%포인트 끌어올렸다. 특히 농산물은 재배면적 감소와 김장철 수요 증가 등으로 11.3% 올랐다.공업제품은 국제유가 하락, 유류세 인하 영향으로 변동이 없었다. 이 중 석유류는 3.5% 하락했다. 전기·수도·가스 물가는 1.2% 상승했다. 서비스 물가는 1.5% 올라 전체 물가를 0.85%포인트 견인했다. 특히 외식, 영화관람, 학원비 등이 포함된 개인서비스 물가가 2.5% 올랐다. 올해 1년간 소비자물가지수는 전년보다 1.5% 상승했다. 지출목적별로는 전년에 비해 음식·숙박(3.1%), 식료품·비주류음료(2.6%), 교통(2.5%), 교육(1.7%), 주택·수도·전기·연료(0.6%) 등이 올랐다 품목별로는 농축수산물은 농산물 등이 올라 3.2%, 공업제품은 석유류 등이 올라 1.3% 각각 상승했다. 전기·수도·가스는 지역난방비 등이 내려 2.9% 하락했다 통계청 관계자는 "폭염 등 기상악화와 국제유가 고공행진 등이 물가 상승을 견인했지만 집세, 전기수도가스 등이 하락하면서 전체 물가 상승세가 둔화했다"고 말했다 구예리기자

Labeling



구예리 기자 유류세 인하 효과 등으로 경기지역 소비자물가 상승률이 5개월 만에 가장 낮은 수준을 기록했다. 경인지방통계청이 31일 발표한 '12월 소비자물가 동향'을 보면 이달 소비자물가지수는 1년 전보다 1.3% 상승했다. 지난 7월 1.2%를 기록한 뒤 5달 만에 가장 낮은 상승률이다. 품목 성질별로 보면 농·축·수산물 5.5% 올라 전체 물가를 0.39%포인트 끌어올렸다. 특히 농산물은 재배면적 감소와 김장철 수요 증가 등으로 11.3% 올랐다.공업제품은 국제유가 하락, 유류세 인하 영향으로 변동이 없었다. 이 중 석유류는 3.5% 하락했다. 전기·수도·가스 물가는 1.2% 상승했다. 서비스 물가는 1.5% 올라 전체 물가를 0.85%포인트 견인했다. 특히 외식, 영화관람, 학원비 등이 포함된 개인서비스 물가가 2.5% 올랐다. 올해 1년간 소비자물가지수는 전년보다 1.5% 상승했다. 지출목적별로는 전년에 비해 음식·숙박(3.1%), 식료품·비주류음료(2.6%), 교통(2.5%), 교육(1.7%), 주택·수도·전기·연료(0.6%) 등이 올랐다. 품목별로는 농축수산물은 농산물 등이 올라 3.2%, 공업제품은 석유류 등이 올라 1.3% 각각 상승했다. 전기·수도·가스는 지역난방비 등이 내려 2.9% 하락했다 통계청 관계자는 "폭염 등 기상악화와 국제유가 고공행진 등이 물가 상승을 견인했지만 집세, 전기수도가스 등이 하락하면서 전체 물가 상승세가 둔화했다"고 말했다 구예리기자

기사

Labeled Data

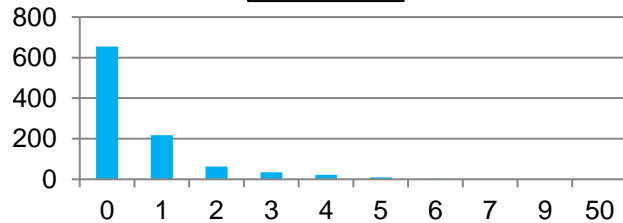
특징	투명 취재원	익명 취재원	전체 서술어	주관적 서술어	미확인 서술어	비문
표현 수	1	1	13	0	0	0

# 데이터 레이블링

## 레이블링

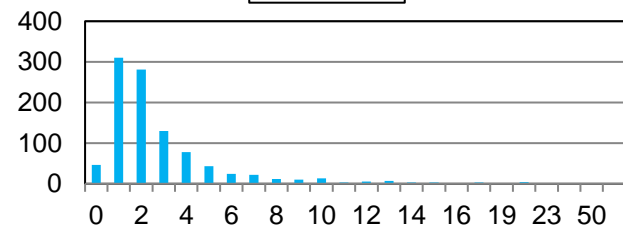
### ➤ 각 변수 데이터 분포 파악

익명 취재원



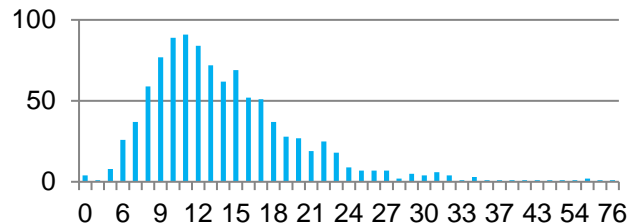
좌측으로 편향 됨 익명 취재원이 많이 집계 되지는 않음

투명 취재원



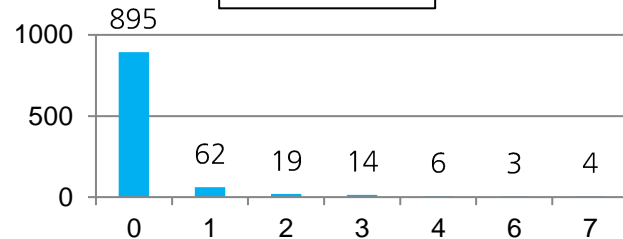
4이하의 데이터가 많은 개수를 보임

서술어 수



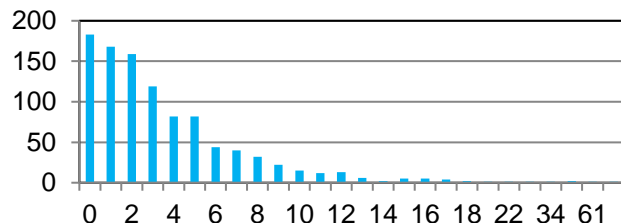
좌측으로 편향된 데이터 분포를 가짐

미확인 서술어 수



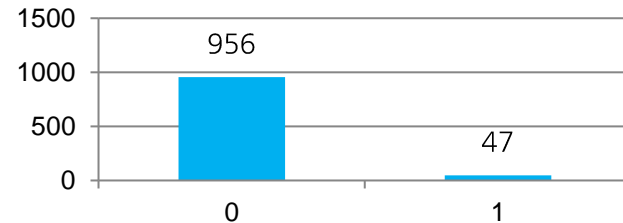
대부분의 데이터가 미확인 서술어를 사용하지 않음

주관적 서술어 수



데이터가 좌측으로 편향되며 0-4 근처에서 많이 누계됨

비문 수



대부분 비문인 것은 없으며 비문인 경우는 비교적 소수

# 데이터 레이블링

## 레이블링

- 최종 예측할 변수를 설정

논문에 의거해 4가지 예측 변수 측정을 위해 가이드라인을 형성. 이후 각 데이터 마다 예측 변수를 계산을 진행. 계산의 정확성을 올리고 속도를 높이기 위해 이진 변수로 최종 예측 변수 변경

### 5-likert-scale



### Binary



기사의 품질을 binary로 분류  
(0 : low quality , 1 : High quality)

#### 최종 예측 변수 공식

1. 투명취재원수 2개 이상
2. 주관적 서술어 수 비율 평균( $\frac{\text{주관적 서술어 수}}{\text{서술어 수}}$ ) 이하
3. 미확인 서술어 수 0개  
위의 모든 조건을 만족하는 경우 1점  
위의 모든 조건을 만족하지 않거나, 비문일 경우 0점

# 데이터 레이블링

## 레이블링

### ➤ 레이블링 데이터 개수 파악

기사번호	매체명	매체유형	매체구분	발행일시	기사카테고리	기사대중소	기사본문글자수	제목	내용
329412764	경기일보	온라인	지역지	2019-01-01 00:00:00	종합	중	1176	105만 고양시	송주현 기자 문
329412768	경기일보	온라인	지역지	2019-01-01 00:00:00	종합	중	1197	고양시 새해에	송주현 기자 문
329412771	경기일보	온라인	지역지	2019-01-01 00:00:00	종합	소	884	안양시, 오는 2	박준상 기자 문
329412773	경기일보	온라인	지역지	2019-01-01 00:00:00	종합	소	759	오명근 의원, 최	최현호 기자 문
329412774	경기일보	온라인	지역지	2019-01-01 00:00:00	종합	소	937	화성시, 2020년	홍완식 기자 문
329412786	경기일보	온라인	지역지	2019-01-01 00:00:00	종합	중	1149	김포대, 중소기업	양형찬 기자 문
329427450	경기일보	온라인	지역지	2019-01-01 00:00:00	종합	소	822	올 신년연휴 안	양광범 기자 문
329437110	경기일보	온라인	지역지	2019-01-01 00:00:00	종합	중	1141	올해부터 수도권	류택현 경기지
329437128	경기일보	온라인	지역지	2019-01-01 00:00:00	종합	중	1094	산업부, 새해부	권혁준 기자 문

데이터 분석에 필요한 새로운 변수 생성

	media	a_num	title	content	w_num	c_source	a_source	s_num	v_num	nf_verb	s_verb	reporter	non_text	s_rate	quality
0	경기일보	329412764	105만 고양	송주현 기자	1176	6	1	18	16	0	8	1	0	50	1
1	경기일보	329412768	고양시 새해	송주현 기자	1197	3	1	14	16	0	6	1	0	37.5	1
2	경기일보	329412771	안양시, 오는 2	박준상 기자	884	2	0	9	10	0	0	1	0	0	1
3	경기일보	329412773	오명근 의원, 최	최현호 기자	759	9	3	9	10	0	2	1	0	20	1
4	경기일보	329412774	화성시, 2020년	홍완식 기자	937	2	0	8	8	0	0	1	0	0	1
5	경기일보	329412786	김포대, 중소기업	양형찬 기자	1149	2	2	9	11	0	0	1	0	0	1
6	경기일보	329427450	올 신년연휴 안	양광범 기자	822	1	2	11	10	1	0	1	0	0	0
7	경기일보	329437110	올해부터 수도권	류택현 경기	1141	3	4	17	10	0	7	1	0	70	0
8	경기일보	329437128	산업부, 새해부	권혁준 기자	1094	3	4	14	17	0	4	1	0	23.529412	1

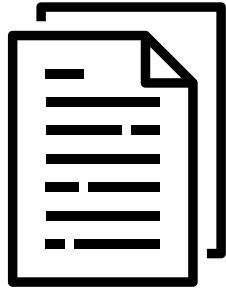
High quality

500

Low quality

503





# 모델 개발



# 모델 개발

## 기존연구조사

- 딥러닝 모델을 적용하기 위해 다양한 논문을 찾아보고 학습하며 적용시킴

논문	저자	발행연도	내용
Credibility Assessment in the News: Do we need to read?	Girgis, Sherry, Eslam Amer, Mahmoud Gadallah	2018	Fake News classification LSTM, GRU
HIERARCHICAL TRANSFORMERS FOR LONG DOCUMENT CLASSIFICATION	Raghavendra Pappagari , Piotr Zelasko, Jesus Villalba, Yishay Carmiel, Najim Dehak	2019	BERT를 활용한 긴 텍스트 문서 분류
딥러닝 사전학습 언어모델 기술 동향	한국전자통신연구원	2020	딥러닝 언어모델인 BERT의 최근 기술 동향을 소개
How to Fine-Tune BERT for Text Classification?	Chi Sun, Xipen Qiu, Yige Xu, Xuanjing Huang	2020	다국어 BERT Model 소개 및 적용

- ✓ 텍스트 분류에 사용되는 자연어 처리 딥러닝 모델을 연구함
- ✓ 긴 글의 기사를 처리하기 위한 방법을 연구하기 위해 다양한 모델을 탐구함
- ✓ 다양한 텍스트 분류 모델을 탐구하였으나, 한국어 기사 분류 목적에 따라 영어로만 학습된 모델은 설계에서 배제함

# 모델 개발

## 적용 모델 소개

### RNN 기반 자연어 모델 - LSTM

- LSTM은 Hochreiter & Schmidhuber (1997)에 의해 소개
- 히든 노드가 방향을 가진 엣지로 연결돼 순환구조를 이루는(directed cycle) 인공신경망의 한 종류
- 역전파 시 기울기가 점차 줄어 학습능력이 저하되는 기울기 소실 문제를 해결하기 위해 개발됨

### GPT2

- 2019년 2월, OPEN AI가 공개한 최신 딥러닝 언어모델
- 기존 transformer 구조에서 encode부분을 제외한, decoder로만 구성된 아키텍처
- RNN구조 대신 positional encoding후에 self-attention을 연산을 적용
- 2천 5백만 이상의 문장으로 학습(wiki + news)한 한국어 모델 KoGPT-2이 있음

### BERT

- 2018년 11월, 구글이 공개한 Transformer 기반 인공지능(AI) 언어 모델 'BERT(Bidirectional Encoder Representations from Transformers)'
- 신문기사와 백과사전 등 23GB의 대용량 텍스트를 대상으로 47억개의 형태소를 사용하여 학습 KorBERT(2019.06, ETRI 개발)이 있음

# 모델 개발

## LSTM 모델 결과

Epoch	Batch size	Learning rate	Optimizer	accuracy
10	64	0.001	Adam	60.4%
20	64	0.0001	Adam	59.4%
30	64	0.00001	Adam	60.4%
10	32	0.001	Adam	55.5%
10	64	0.001	Adam	58.4%
20	128	0.001	Rmsprop	58.4%
10	32	0.001	Adam	60.4%
10	64	0.001	Adam	58.4%
20	32	0.001	Rmsprop	57.4%

[단방향 LSTM] 성능

[LSTM+attention] 성능

[영화 리뷰 감성분석 데이터의 embedding 값 초기 가중치 + Bidirectional LSTM] 성능

- ✓ LSTM을 사용해 모델링을 진행한 결과, 단방향 LSTM을 사용한 경우와 전이학습 시킨 데이터의 가중치를 초기 가중치로 둔 LSTM 모델을 사용한 경우 가장 높은 성능을 나타냄

# 모델 개발

## 최종 Model 설계 input 설정

- ① BERT 최대 Input크기인 512에 맞게, ETRI에서 제공한 형태소 분석기를 통해 레이블 기준에 근거하여 불필요한 어절을 제거한 뒤, '명사+지정사(이다)', '동사구', '주격' 추출

### 기사 예시

경기도의 중심도시인 고양시가 되기까지 이들은 후배들에게 자리를 물려주고 인생 2막을 준비 중이다.

### 불필요한 어절 제거

고양시가 되기까지 이들은 물려주고 인생 준비 중이다.

### 필요한 어절('명사+지정사(이다)', '동사구', '주격') 및 품사 추출

고양시/NNP 가/ JKS 되/VV 기/ETN 까지/JX 이/NP 들/XSN 은/JX 물려주/VV 고/EC 중/NNB 이/VCP 다/EF ./SF

### 토큰화

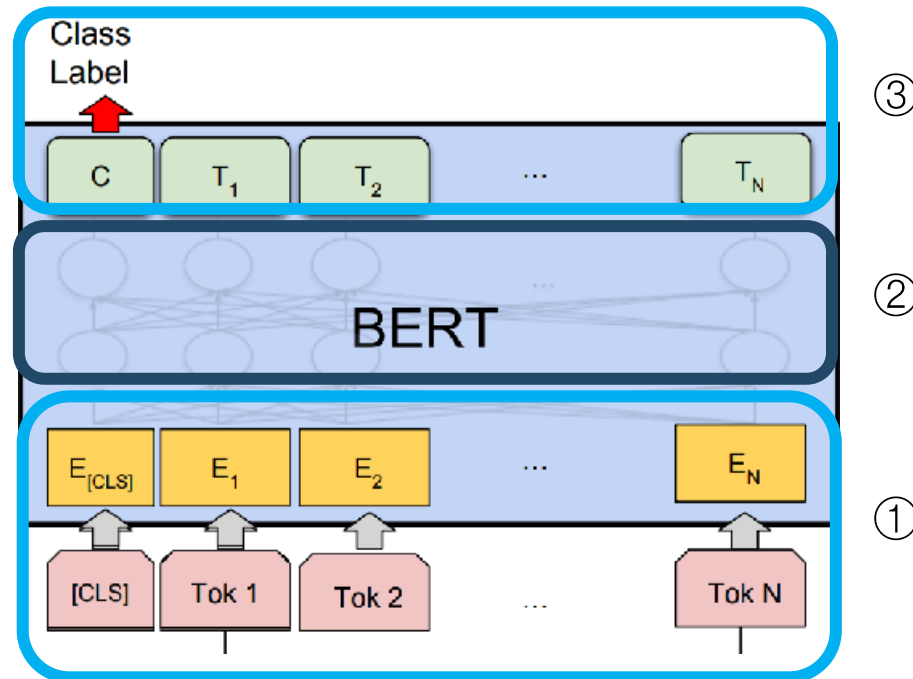
['[CLS]', '고양시/NNP\_', '가/JKS', '되/VV\_', '기/ETN\_', '까지/JX\_', '이/NP\_', '들/XSN\_', '은/JX\_', '물', '려주/VV\_', '고/EC\_', '중/NNB\_', '이/VCP\_', '다/EF\_', './SF\_', '[SEP]']

\*BERT의 input 형태에 맞게 문장 앞 뒤에 각각 [CLS]와 [SEP]을 추가

# 모델 개발

## 최종 Model 설계

- ② Bert의 overfitting을 방지하기 위해 총 12개의 layer 중 마지막 층을 제외하고 모두 freeze 시킨 후 전이학습 진행
- ③ 기존 Bert의 output층과 연결되어 있는 기사 분류 모델에 불필요한 MLM(Masked Language Model), NSP(Next Sentence Prediction)층을 제거하는 대신, Binary classification을 위해 Bert Layer의 output값을 Sigmoid Dense 층에 학습



# 모델 개발

## KoGPT2, KorBERT

### Kogpt2 모델 결과

Epoch	Batch Size	Learning rate	Optimizer	Accuracy
5	8	5e-5	AdamW	45%
10	8	3e-5	Rectified Adam	48%
5	16	3e-5	Rectified Adam	50%

- ✓ Epoch 5, Batch Size 16, Learning rate 3e-3, Optimizer 는 Radam에서 50%정도의 정확도를 보임.

### Korbert 모델 결과

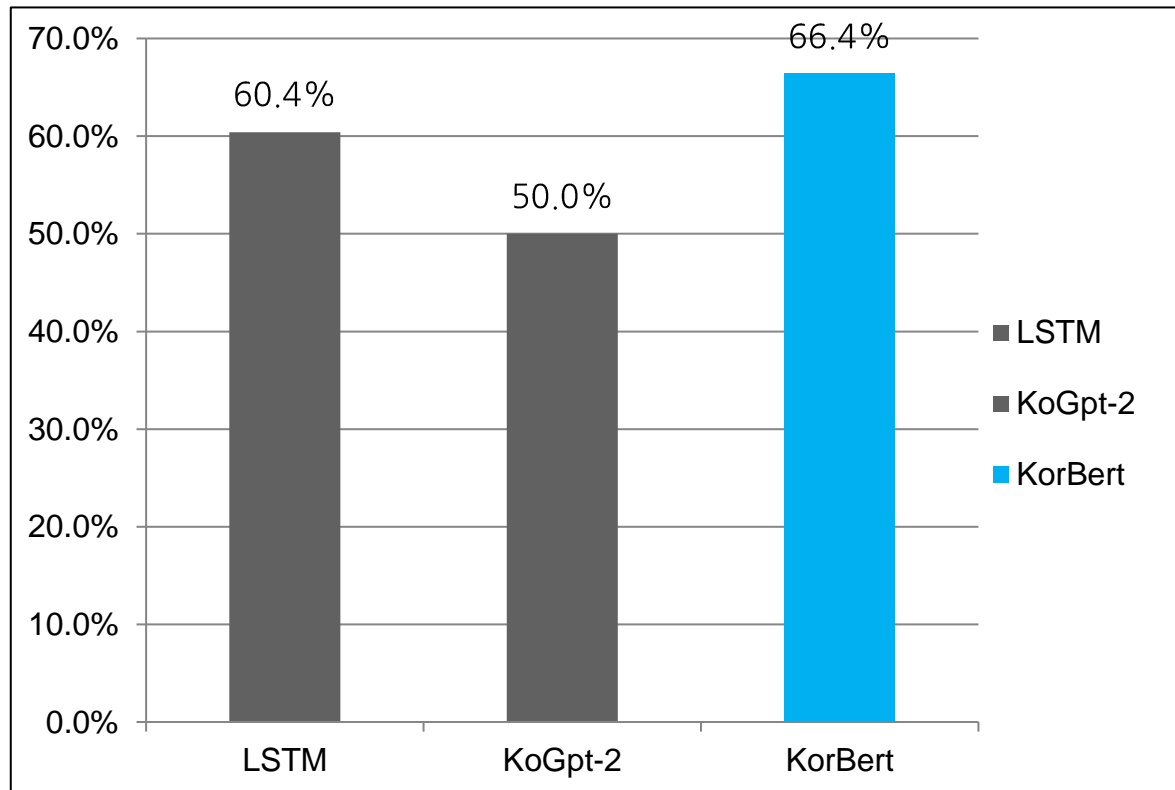
Epoch	Batch Size	Learning rate	Optimizer	Accuracy
10	16	3e-5	Rectified Adam	66.4%
10	8	3e-5	Rectified Adam	66.1%
5	32	3e-5	Adam	65.4%
5	32	3e-5	Rectified Adam	65.7%

- ✓ Epoch 10, Batch Size 16, Learning rate 3e-5, Optimizer 는 Radam에서 66.4%정도의 정확도를 보임

# 모델 개발

## 실험 결과

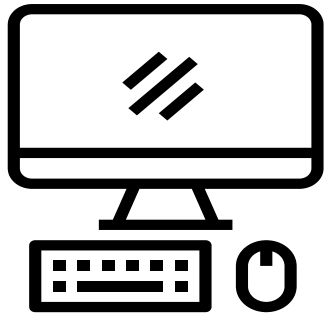
- 각 모델들을 비교 할 때 KorBert에서 안정적이고 높은 accuracy를 가짐



\*  $Train\_data : Test\_data = 9 : 1$

\*  $Accuracy = 100 \times \text{Correct \# of Test data} / \text{All \# of Test data}$





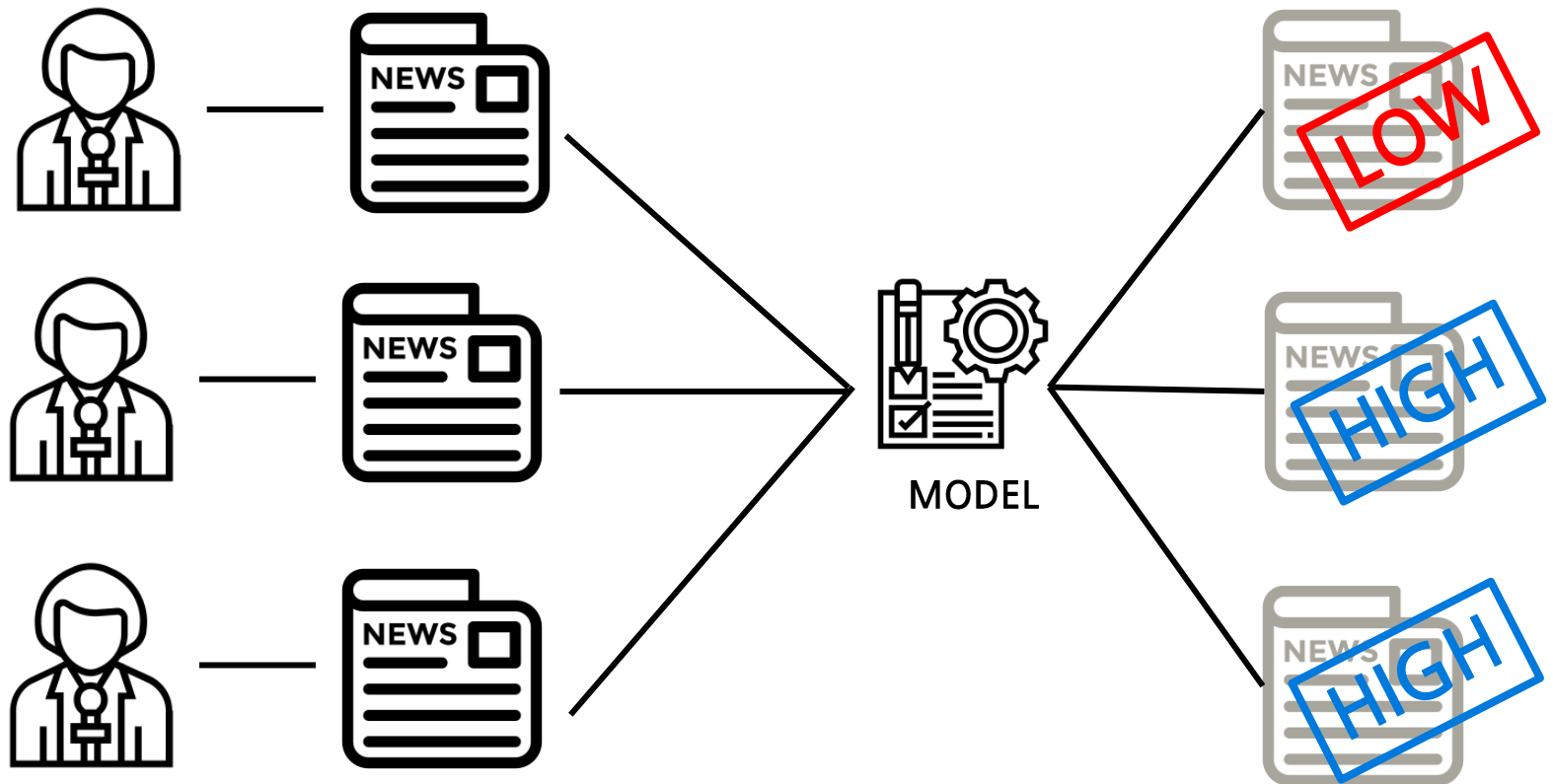
# 기대효과 및 웹 서비스

---

# 기대 효과 및 개선 방향

## 기대 효과 & 활용 방안

- 기사 작성 직군의 기사 품질 측정 서비스 제공
- 궁극적으로 다양한 기사들의 전반적 품질 향상 기대



---

*Thank You*

*&*

*Q&A*

# 부록

## 선정 기준

### • 품질 기준

- **미확인 서술어** : ‘~로 전해졌다. ~로 알려졌다. ~라는 주장이 제기됐다. ~라는 의혹이 제기됐다, ~인 것으로 알려져 파문이 예상된다
- **주관적 서술어** : 강조했다, 주장했다, 당부했다. 아쉬워했다. 우려했다, 단언했다, 촉구했다, 경고했다, 선언했다, 지적했다, 비판했다, 설명했다, 해명했다, 반문했다, 의혹을 제기했다, 의아해했다, 시사했다, 기대했다, 전망했다, 평가했다, 토론했다, 귀띔했다
- [~라는 분석이다(또는 ~로 분석되고 있다, 이하 마찬가지), 지적이다(지적을 받고 있다), 비판이다(비판이 나오고 있다), 생각이다, 계획이다, 주장이다(주장이 나오고 있다), 자평이다, 확신이다, 의문이다, 소식이다, 평가다(평가도 나온다), 설명이다, 입장이다, 방침이다, 분위기이다, 우려가나오고 있다, 가능성도 거론되고 있다, 목소리가 높아지고 있다, 의혹을 사고있다, 의문이 제기된다, 격찬이 쏟아졌다], [~로 보인다, 관측되고 있다, 전망되고 있다, 예상된다, 추정된다, 알려졌다, 전해졌다], [~를 심각하게 한다, ~한셈이다, ~해야 할 판이다] 등이다. ‘~라는 겁니다’ 식의 표현도 여기에 포함된다.
- **개인 투명 취재원**: 사람의 성과 이름이 명확히 제시된 경우(예, 홍길동은~)
- **단체 투명 취재원**: 기관이나 단체의 이름이 명확히 제시된 경우이며 집합명사 취재원도 여기에 포함했다. 단체 취재원은 (공식적으로) 알려진 내용 또는(공식적인) 발표내용을 인용하는 경우에 주로 쓰인다. 단체취재원은 대체로 실명이다(예, 경찰청은~, 경찰은~, 청와대는~, 한나라당은~, 전교조는~).
- **자료 투명 취재원**: 자료나 문건의 이름이 명확히 제시된 경우다(예, 000의조사에 따르면~, A씨는 법원에 낸 소장에서~, 김 씨의 고발장에는~, 000자료를 보면~).
- **익명 개인 취재원** (예, 김 모 씨는~, 박 아무개는~, A씨는~, 서울시의 한 고위간부는~, 경제부처 모 서기관은~, 이름을 밝히기를 거부한 한 당직자는~)
- **불특정 다수 취재원**: 성명이 밝혀져 있지 않은 복수의 인물 취재원(예, 주민들은~, 일부 교사들은~)

# 부록

## 출처

---

- 템플릿 출처 : 조땅의 템플릿
- 논문 출처
  - 취재원 사용의 원칙과 현실 : 세월호 보도를 중심으로(송상근)
  - 뉴스평가지수의 개발과 적용(박재영,이완수,2010)
  - 방송뉴스의 언어와 표현 - 뉴스언어의 객관성과 공정성을 중심으로(박재영,이완수, 2010)
  - 방송 법조뉴스의 품질 연구 정확성,심층성, 불편부당성을 중심으로 ( 박성호,윤형민, 2016)
  - 방송뉴스의 단독 보도 품질 (유수정, 오건호 ,2020)
  - Credibility Assessment in the News: Do we need to read?(Girgis, Sherry, Eslam Amer, Mahmoud Gadallah,2018)
  - HIERARCHICAL TRANSFORMERS FOR LONG DOCUMENT CLASSIFICATION(Raghavendra Pappagari , Piotr Zelasko, Jesus Villalba, Yishay Carmiel, Najim Dehak, 2019)
  - 딥러닝 사전학습 언어모델 기술 동향 (한국전자통신연구원 , 2020)
  - How to Fine-Tune BERT for Text Classification?(Chi Sun, Xipen Qiu, Yige Xu, Xuanjing Huang,2020)
- 픽토그램 : nounproject