

# korpatbert 분류모델 실행방법 (23.01.21)

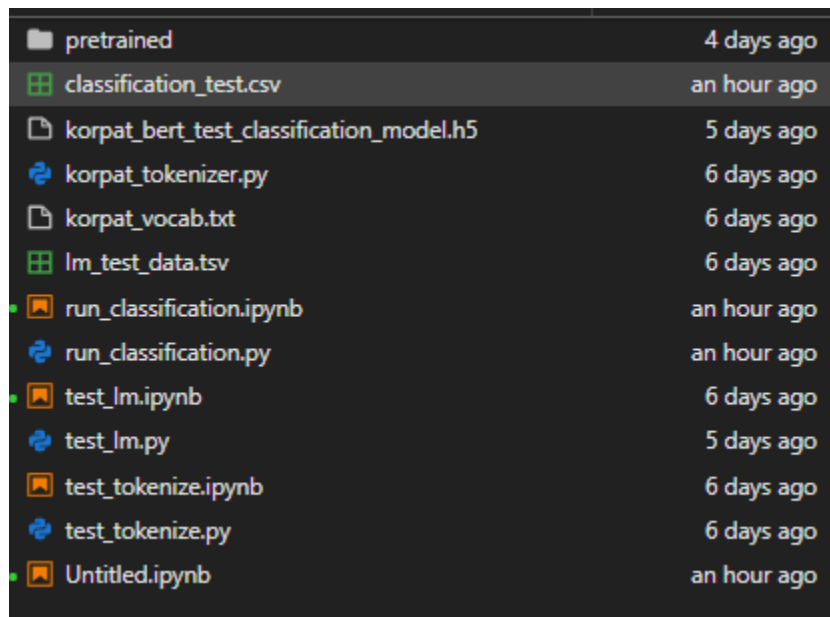
1. 모델을 돌리기 위해 필요한 라이브러리를 설치함.

[ 사전 설치 요구사항 목록]

프로그래밍	버전	설치 안내 경로	필수
python	3.6 이상	<a href="https://www.python.org/">https://www.python.org/</a>	Y
anaconda	4.6.8 이상	<a href="https://www.anaconda.com/">https://www.anaconda.com/</a>	N
tensorflow	2.2.0 이상	<a href="https://www.tensorflow.org/install/pip?hl=ko">https://www.tensorflow.org/install/pip?hl=ko</a>	Y
sentencepiece	0.1.96 이상	<a href="https://github.com/google/sentencepiece">https://github.com/google/sentencepiece</a>	N
mecab-ko	0.996-ko-0.0.2	<a href="https://bitbucket.org/eunjeon/mecab-ko-dic/src/master/">https://bitbucket.org/eunjeon/mecab-ko-dic/src/master/</a>	Y
mecab-ko-dic	2.1.1		Y
mecab-python	0.996-ko-0.9.2		Y
python-mecab-ko	1.0.11 이상	<a href="https://pypi.org/project/python-mecab-ko/">https://pypi.org/project/python-mecab-ko/</a>	Y
keras	2.4.3 이상	<a href="https://github.com/keras-team/keras">https://github.com/keras-team/keras</a>	N
bert_for_tf2	0.14.4 이상	<a href="https://github.com/kpe/bert-for-tf2">https://github.com/kpe/bert-for-tf2</a>	N
tqdm	4.59.0 이상	<a href="https://github.com/tqdm/tqdm">https://github.com/tqdm/tqdm</a>	N
soynlp	0.0.493 이상	<a href="https://github.com/lovit/soynlp">https://github.com/lovit/soynlp</a>	N

2.모델 폴더안에 있는 classification\_text.csv 파일을 연다.

2.A열에 분류하고자 하는 내용을 삽입한다(A열(sentence)과 B열(label)의 칼럼명은 그대로)



	A	수식 입력줄	C
1	sentence	label	
2	"1. 서론		
3	"1. 머리말		
4	"1. 연구의 배경과 목적		
5	"1. 연구의 필요성		
6	"통합		
7	"I. 서론		
8	"통합		
9	"1. 서론		
10	"1. 서론		
11			
12			
13			
14			
15			
16			

## 4.run\_classification.py 파일을 실행함

```
irteam@2fdb20d1b72b:~/home/korpatbert_v1.0$ python run_classification.py
```

```
irteam@2fdb20d1b72b:~/home/korpatbert_v1.0$ python run_classification.py
2023-01-20 17:02:16.680066: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic library libcuda.so.1
2023-01-20 17:02:16.718347: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1561] Found device 0 with properties:
pciBusID: 0000:1a:00.0 name: GeForce RTX 2080 Ti computeCapability: 7.5
coreClock: 1.545GHz coreCount: 68 deviceMemorySize: 10.76GiB deviceMemoryBandwidth: 573.69GiB/s
2023-01-20 17:02:16.719265: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1561] Found device 1 with properties:
pciBusID: 0000:1b:00.0 name: GeForce RTX 2080 Ti computeCapability: 7.5
coreClock: 1.545GHz coreCount: 68 deviceMemorySize: 10.76GiB deviceMemoryBandwidth: 573.69GiB/s
2023-01-20 17:02:16.720146: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1561] Found device 2 with properties:
pciBusID: 0000:60:00.0 name: GeForce RTX 2080 Ti computeCapability: 7.5
coreClock: 1.545GHz coreCount: 68 deviceMemorySize: 10.76GiB deviceMemoryBandwidth: 573.69GiB/s
2023-01-20 17:02:16.720362: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic library libcudart.so.10.1
2023-01-20 17:02:16.722181: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic library libcublas.so.10
2023-01-20 17:02:16.723838: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic library libcufft.so.10
2023-01-20 17:02:16.724144: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic library libcudnn.so.7
2023-01-20 17:02:16.726089: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic library libcusolver.so.10
2023-01-20 17:02:16.727171: I tensorflow/stream_executor/platform/default/dso_loader.cc:44] Successfully opened dynamic library libcusparse.so.10
2023-01-20 17:02:16.727453: W tensorflow/stream_executor/platform/default/dso_loader.cc:55] Could not load dynamic library 'libcudnn.so.7'; dlopen error: libcudnn.so.7: cannot open shared object file: No such file or directory
2023-01-20 17:02:16.727467: W tensorflow/core/common_runtime/gpu/gpu_device.cc:1598] Cannot dlopen some GPU libraries. Please make sure the missing libraries mentioned above are installed properly if you would like to use GPU. Follow the
guide at https://www.tensorflow.org/install/gpu for how to download and setup the required libraries for your platform.
Skipping registering GPU devices...
2023-01-20 17:02:16.727763: I tensorflow/core/platform/cpu_feature_guard.cc:143] Your CPU supports instructions that this TensorFlow binary was not compiled to use: AVX2 AVX512F FMA
2023-01-20 17:02:16.738286: I tensorflow/core/platform/profile_utils/cpu_utils.cc:102] CPU Frequency: 2700000000 Hz
2023-01-20 17:02:16.740292: I tensorflow/compiler/xla/service/service.cc:168] XLA service 0x7f9f28000b20 initialized for platform Host (this does not guarantee that XLA will be used). Devices:
2023-01-20 17:02:16.740331: I tensorflow/compiler/xla/service/service.cc:176] StreamExecutor device (0): Host, Default Version
2023-01-20 17:02:16.741944: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1102] Device Interconnect StreamExecutor with strength 1 edge matrix:
2023-01-20 17:02:16.741967: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1108]

==== 평가데이터 샘플 출력 및 전처리 시작 ====
      sentence  label
0  "1. 서론\n\n차물들의 형상은 사물들을 거의 간직하고 있지 않다. 대상을 채...   NaN
1  "1. 머리말\n\n인간의 커뮤니케이션은 처음에는 음성과 몸짓 같은 자연적이고 직접...   NaN
2  "1. 연구의 배경과 목적\n\n국내의 게임산업은 2006년 발생했던 '바다이야기'...   NaN
3  "1. 연구의 필요성\n\n관객들은 만성 퇴행성 질환 중 가장 흔한 질환 중 하나로...   NaN
4  "통합\n\n36 1. 작품개요 프로젝트 맵핑은 일반 피사체, 건물, 3차원 형태...   NaN
5  "1. 서론\n\n우리나라도 기술의 중요성이 높아지는 만큼 실제 시장에서도 핵심기술...   NaN
6  "통합\n\n핵에 대한 중요성이 급격히 부각되는 상황이지만 국 내에서는 대외수를 연...   NaN
7  "1. 서론\n\n최근 평균 수명이 증가되면서 고령인구도 점차증가하고 있다. 20...   NaN
8  "1. 서론\n\n최근 온라인 구매 비율이 60%를 넘어서고 있고, 코로나19 상황...   NaN
데이터 전처리 진행 중: 1it [00:00, 3.53it/s]
```

## 5. 모델폴더에 생성된 classification\_results.csv을 열어 결과 확인함.

```

===> 평가데이터 샘플 출력 및 전처리 시작 <===
      sentence label
0 "1. 서 론¶¶¶¶¶ 사물들의 형상은 사물들을 거의 간직하고 있지 않다. 대상을 재... NaN
1 "1. 머리말¶¶¶¶¶ 인간의 커뮤니케이션은 처음에는 음성과 몸짓 같은 자연적이고 직접... NaN
2 "1. 연구의 배경과 목적¶¶¶¶¶ 국내의 게임산업은 2006년 발생했던 '바다이야기'... NaN
3 "1. 연구의 필요성¶¶¶¶¶ 관절염은 만성 퇴행성 질환 중 가장 흔한 질환 중 하나로... NaN
4 "통합¶¶¶¶¶ 36 1. 작품개요 프로젝션 맵핑은 일반 피사체, 건물, 3차원 형태... NaN
5 "1. 서론¶¶¶¶¶ 우리나라도 기술의 중요성이 높아지는 만큼 실제 시장에서도 핵심기술... NaN
6 "통합¶¶¶¶¶ 책에 대한 중요성이 급격히 부각되는 상황이지만 국 내에서는 내외수를 연... NaN
7 "1. 서 론¶¶¶¶¶ 최근 평균 수명이 증가되면서 고령인구도 점차증가하고 있다. 20... NaN
8 "1. 서론¶¶¶¶¶ 최근 온라인 구매 비율이 60%를 넘어서고 있고, 코로나19 상황... NaN
데이터 전처리 진행중: 9it [00:03, 2.30it/s]
test_x [[ 5 242 19 ... 12 2945 6]
 [ 5 242 19 ... 45 773 6]
 [ 5 242 19 ... 2270 5427 6]
 ...
 [ 5 242 1924 ... 68 6379 6]
 [ 5 242 19 ... 25 125 6]
 [ 5 242 19 ... 1396 8098 6]]
test_x 9
text_x 예측embedding [[9.9838364e-01 1.5832724e-03 3.3072927e-05]
 [9.2701745e-01 7.2733551e-02 2.4893423e-04]
 [9.1560465e-01 8.4329225e-02 6.6116154e-05]
 [9.9979800e-01 1.6593150e-04 3.6148274e-05]
 [9.4496036e-01 5.4980241e-02 5.9334867e-05]
 [9.9800259e-01 1.9525884e-03 4.4812183e-05]
 [9.9643731e-01 3.5227502e-03 3.9911822e-05]
 [9.8443168e-01 1.5482396e-02 8.5874621e-05]
 [9.9986184e-01 1.1996670e-04 1.8263305e-05]]
예측값 [0 0 0 0 0 0 0 0 0]
text_x 예측개수 9

===> 분류가 완료 되었습니다

```

	A	B	C	
1		원문sender	예측값	
2	0	"1. 서 론	0	
3	1	"1. 머리말	0	
4	2	"1. 연구의	0	
5	3	"1. 연구의	0	
6	4	"통합	0	
7	5	"I. 서론	0	
8	6	"통합	0	
9	7	"1. 서 론	0	
10	8	"1. 서론	0	
11				

분류결과파일예시(0과1로분류)  
(classification\_results.csv)