DLthon

협박, 갈취, 직장 내 괴롭힘, 기타 괴롭힘
4가지 대화 유형 Class를 분류하는 딥러닝 모델 구축 Project

Team name : Ttol mang

Team members : 김지원, 서민성, 임정훈, 류의성

Index

Problem definition

TUNiB에서 제공하는 DKTC 훈련 데이터셋을 활용하여
협박, 갈취, 직장 내 괴롭힘, 기타 괴롭힘 4가지 대화 유형 Class를 분류하는 딥러닝 모델 구축

TUNiB

클래스	Class No.	# Training	# Test
협박	00	896	100
갈취	01	981	100
직장 내 괴롭힘	02	979	100
기타 괴롭힘	03	1,094	100
~~일반~~	~~04~~		~~100~~

[DKTC training data set]

▶

4가지 유형 Class 분류

EDA

Train data set의 경우 줄 바꿈 기호인 '₩n'가 있어
해당 데이터 제거 및 2개 이상의 연속된 공백 제거 진행

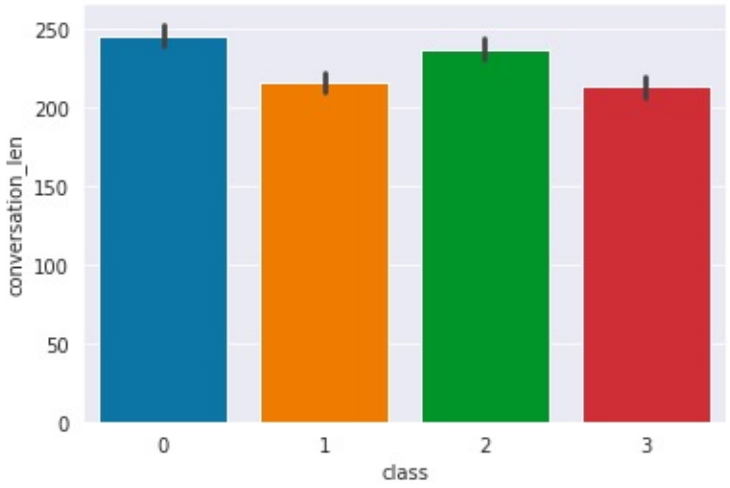
'35번 손님 아이스커피 두잔나왔습니다\n아이스커피? \n네 맛있게드세요\n저기요 아가씨 나는 아아스 시킨적이 없는데?\n아 분명 오늘 날이 더우시다고 아이스로 시키셨는데요\n내가 그랫어? \n네분명히.\n아그런 기억이 없는데? 아가씨가잘못안거 아니야?\n아니요. 오늘 손님이 첫 주문이라 확실히 기억하고 있습니다\n아가씨. 왜이렇게 유도리가 없이 굴어 그냥 아 제가 잘못 주문 받았습니다 하면 되지?\n.네?.\n어휴 유도리 없어 그냥 마실게'

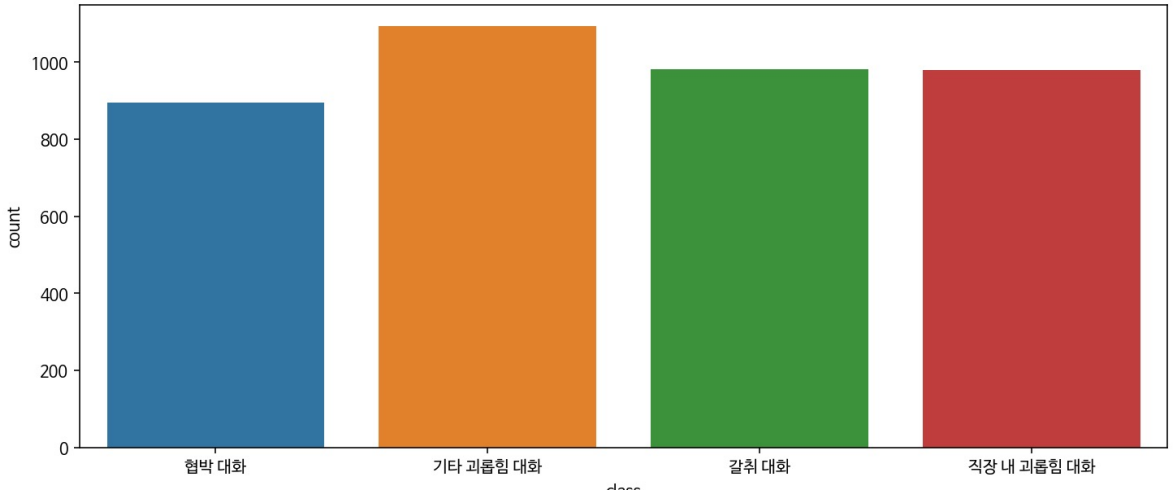
▼

'35번 손님 아이스커피 두잔나왔습니다 아이스커피 네 맛있게드세요 저기요 아가씨 나는 아아스 시킨적이 없는데 아 분명 오늘 날이 더우시다고 아이스로 시키셨는데요 내가 그랫어 네분명히 아그런 기억이 없는데 아가씨가잘못안거 아니야 아니요 오늘 손님이 첫 주문이라 확실히 기억하고 있습니다 아가씨 왜이렇게 유도리가 없이 굴어 그냥 아 제가 잘못 주문 받았습니다 하면 되지 네 어휴 유도리 없어 그냥 마실게'

EDA

데이터 셋의 클래스 별 데이터는 기타 괴롭힘 1,094건, 갈취 981건, 직장 내 괴롭힘 979건, 협박 896건이며,
데이터 셋의 클래스 별 개수는 다음과 같고, “기타 괴롭힘"이 조금 많은 것으로 확인됨.



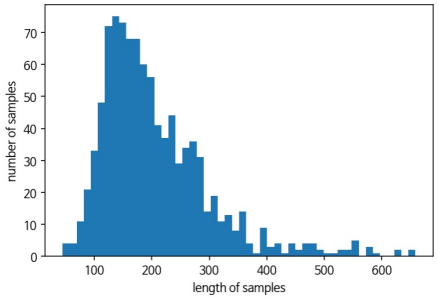


[각 클래스별 문장 길이 분포]

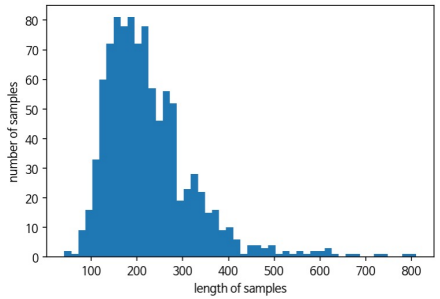
EDA

각 클래스 별 문장 길이 분포는 왼쪽으로 편중되어 있으며, “0”에 해당되는 협박이 약간 길지만 평균 길이는 199~234 사이로 큰 차이를 보이지 않음

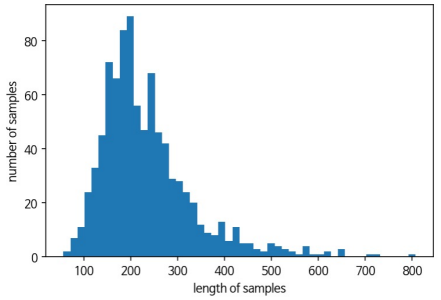
갈취 대화 의 최대 길이 :658
갈취 대화 의 평균 길이 :205.47604485219165



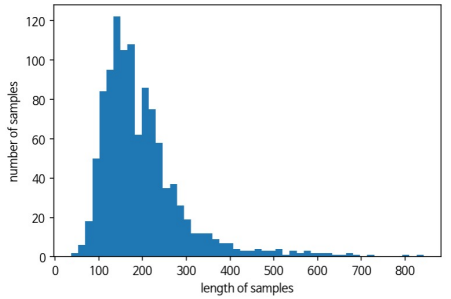
직장 내 괴롭힘 대화 의 최대 길이 :810
직장 내 괴롭힘 대화 의 평균 길이 :226.62819203268643



협박 대화 의 최대 길이 :807
협박 대화 의 평균 길이 :234.85714285714286



기타 괴롭힘 대화 의 최대 길이 :842
기타 괴롭힘 대화 의 평균 길이 :199.94149908592323



[각 클래스별 문장 길이 분포]

EDA - Data Augmentation

각 클래스별 평균 데이터는 987.5건으로 학습을 진행함에 있어 적은 데이터로 판단되어
Back Translation Data Augmentation 을 진행하여 비교하고자 함

원문
데이터

안녕하세요 > Hello > 하이

원어를 타국어로 번역 후 재 번역을 통한 Data Augmentation 기법

안녕하세요 > こんにちは。 > 안녕

원문
데이터

[Back Translation 예시]

EDA - Data Augmentation

Google Translation을 사용하여 Augmentation을 진행하였지만
진행결과 원문의 의미 및 문맥이 제대로 전환이 안되는 것을 확인

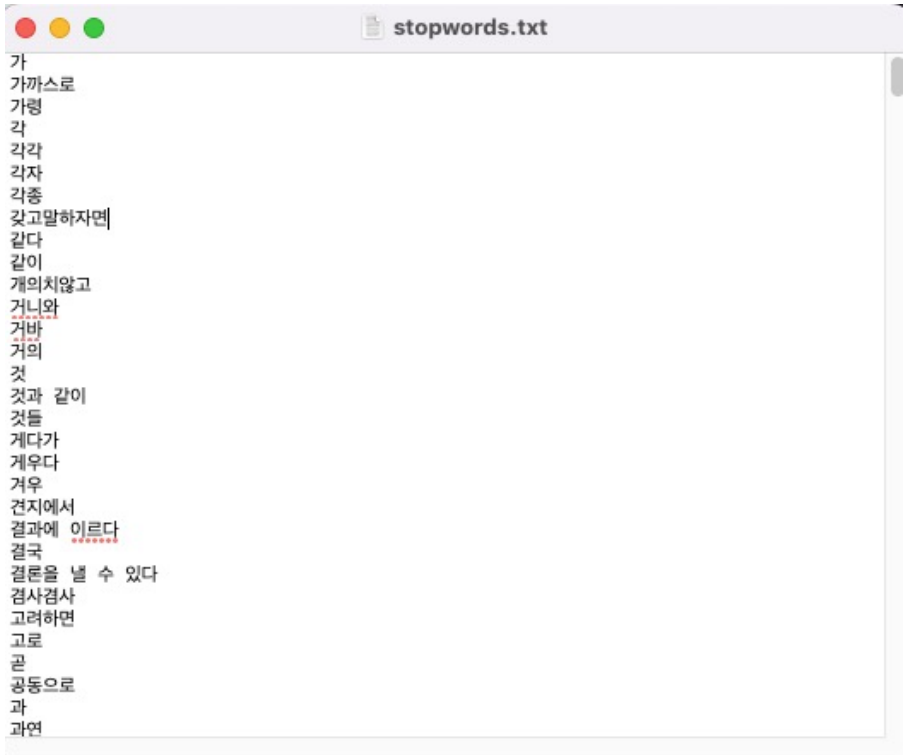
(원문) 네네 무슨 일 때문에 전화주셨나요 우리 애가 지우개 하나 훔친거 가지고 애들 앞에서 면박줬니 그런게 아니고 댁 내 따님분이 한 두번 훔친것도 아니고 여러번 훔친 게 확인 되어서 경찰부르는건 너무 심할거 같아서 꾸중 한 마디 한겁니다 아니 그 구멍가게 물건이 얼마 한다고 그러냐고 내 애가 맨날 훔친거 확실해 네 본 애들도 있고 확실합니다 됐고 내 애 울면서 하교 했어 이번 일 각오하는 게 좋을거야 너도 애 있다며 제 애까지 끌어들이지 마세요 내 애도 울면서 들어왔는데 그냥 못 넘어가 니 애 죽여버릴거야 이러지 마세요 저도 곤란합니다 단지 교육 차원에서 한 마디 했을 뿐이라고요 됐고 아 저기 네 애 보이네 죽여버릴테니까 니 아들 죽은거 보고 그때 반성해 남 애 함부로 울리면 어떻게 되는지

(일어) 직업을 위해 무엇을 부르셨나요?내 아이는 지우개 앞에서 한 지우개를 훔쳐서 파산했습니다.나는 내 아이가 항상 훔치고 있다고 확신합니다.나는 4 명의 자녀가 있고 확실합니다.'확실히 아이를 울고 거부했습니다.이것을 준비하는 것이 좋습니다.나는 그것을 버리지 않습니다.나는 또한 이것을 하기가 어려웠다.

(영어)직업을 위해 무엇을 부르셨나요?내 아이는 한 번의 지우개를 훔치고 아이들 앞에서 그것을 박살 냈습니다. 나는 내 아이가 항상 훔치고 있다고 확신합니다.나는 4 명의 아이들이 있고 확실합니다. 그리고 나는 아이들을 울고 해산했다고 확신합니다.이것에 대비하는 것이 좋습니다. 나는 그것을 버리지 않을 것입니다.나도 이것도 하기가 어려웠다.

Tokenization and Embedding - Stopword

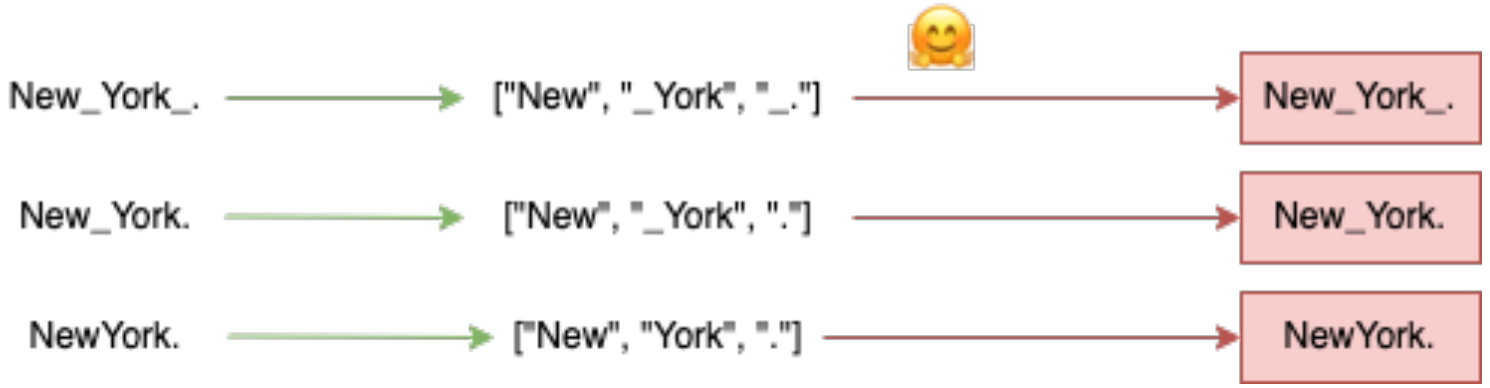
불용어의 경우 595개의 사전을 구축하여 제거 진행



[Stopword list]

https://deep.chulgil.me/hangugeo-bulyongeo-riseuteu/

DL(CNN, LSTM, Bi_LSTM)의 Tokenization의 경우
OOV 문제를 개선하고자 Sentencepiece 를 사용하여 Tokenization을 진행



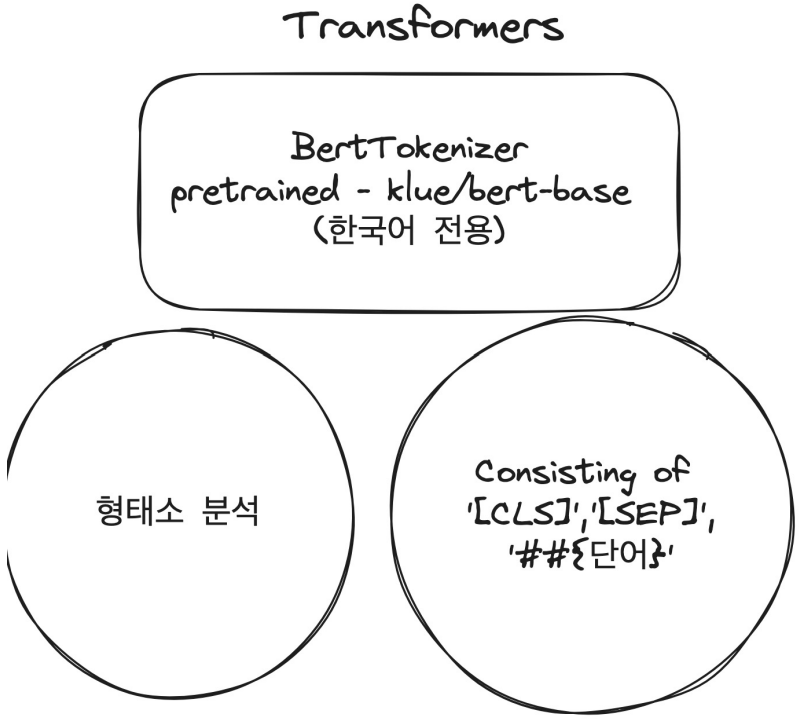
[Sentencepiece 예시]

```
[149, 551, 3082, 34, 611, 28, 574, 16355, 602, 627, 2346, 3586, 26, 3586, 2076, 7, 393, 634, 16534, 31, 6708, 149,
2654, 9224, 1055, 299, 51, 34, 926, 926, 3120, 3586, 2076, 551, 1744, 19, 89, 64, 7895, 6708, 160, 25, 14880, 160,
144, 352, 19, 1057, 5475, 3586]
['_나는', '_당신이', '_스스로', '를', '_죽', '_이', '도록', '_격', '려', '하지', '_않습니다', '.', '_죄송합니다', '.', '_만약', '_
내가', '_혼자', '_죽을', '_것이라', '면', ',', '_나는', '_혼자서', '_죽었', '_기', '_때문에', '_우리', '를', '_죽이고', '_죽이고',
'_싶습니다', '.', '_만약', '_당신이', '_그것', '을', '_할', '_수', '_없다면', ',', '_당신', '은', '_당신과', '_당신', '의', '_가
족', '을', '_죽일', '_것입니다', '.']
나는 당신이 스스로를 죽 이도록 격려하지 않습니다. 죄송합니다. 만약 내가 혼자 죽을 것이라면, 나는 혼자서 죽었 기 때문에 우리를 죽이고 죽이고 싶습니다.
만약 당신이 그것을 할 수 없다면, 당신은 당신과 당신의 가족을 죽일 것입니다.
```

[Sentencepiece 실행 결과]

Tokenization and Embedding

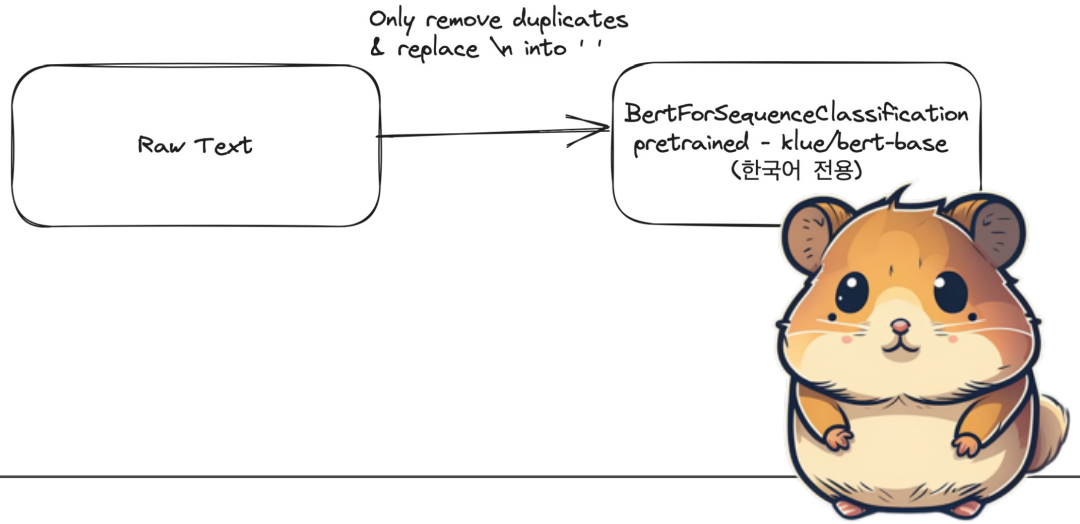
BERT Tokenizer는 '[CLS]'와 '[SEP]' 같은 special tokens을 줌으로써
attention weight가 각 위치마다 생길 수 있도록 도와준다.



'[CLS] 지금 너 스스로를 죽여달라고 애원하는 것인가? [SEP] 아닙니다. 죄송합니다. [SEP] 죽을 거면 혼자 죽지 우리까지 사건에 휘말리게 해? 진짜 죽여 버리고 싶게. [SEP] 정말 잘못했습니다. [SEP] 너가 선택해. 너가 죽을래 네 가족을 죽여줄까. [SEP] 죄송합니다. 정말 잘못했습니다. [SEP] 너에게는 선택권이 없어. 선택 못한다면 너와 네 가족까지 모조리 죽여버릴거야. [SEP] 선택 못하겠습니다. 한번만 도와주세요. [SEP] 그냥 다 죽여버려야겠군. 이의 없지? [SEP] 제발 도와주세요. [SEP]'

[CLS/SEP 예시]

Minimal Pre-processing



Tokenization and Embedding

ML 모델링을 위한 공백기반 토큰화 및 TF-IDF진행
이후 각각의 혼동행렬을 통한 성능비교 예정

공백기반 토큰화

TF-IDF

Modelling

Modelling의 경우 기본 Linear SVC, LSTM, Bi_LSTM, CNN, BERT를 진행하고
추가적으로 LSTM, Bi_LSTM의 경우 Epoc 마다 Back Translation을 진행하여 Augmentation 진행

Linear SVC

BERT

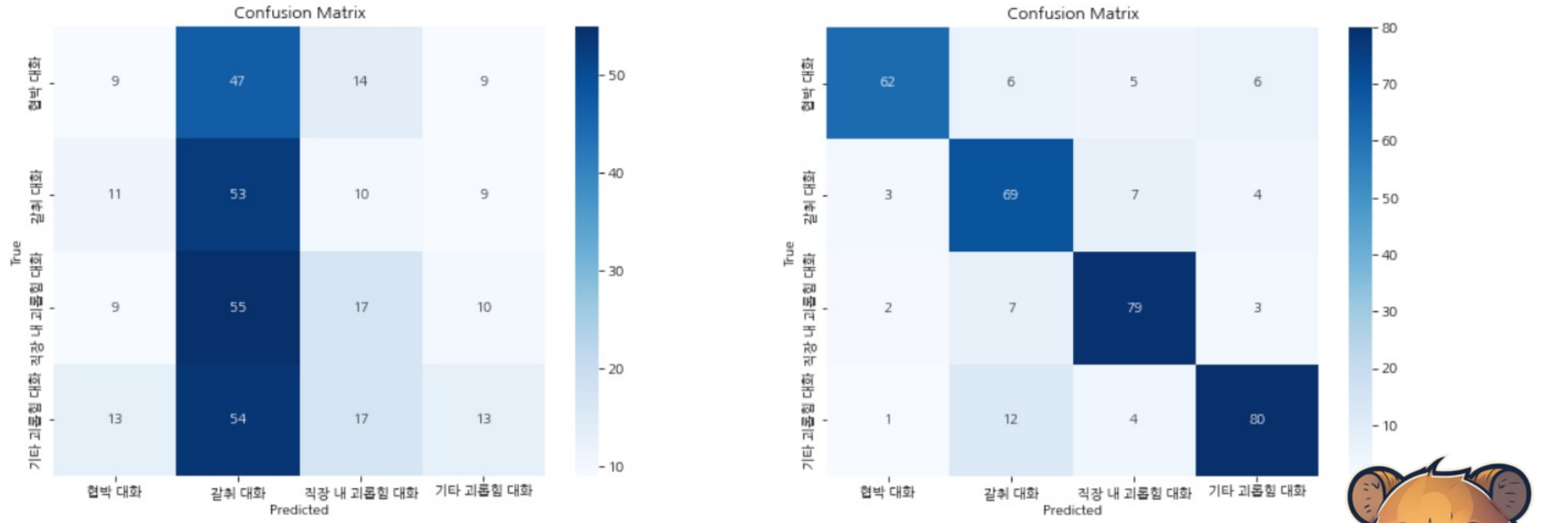
CNN
LSTM
Bi_LSTM

+

BT_LSTM
BT_Bi_LSTM

Modelling - Linear SVC

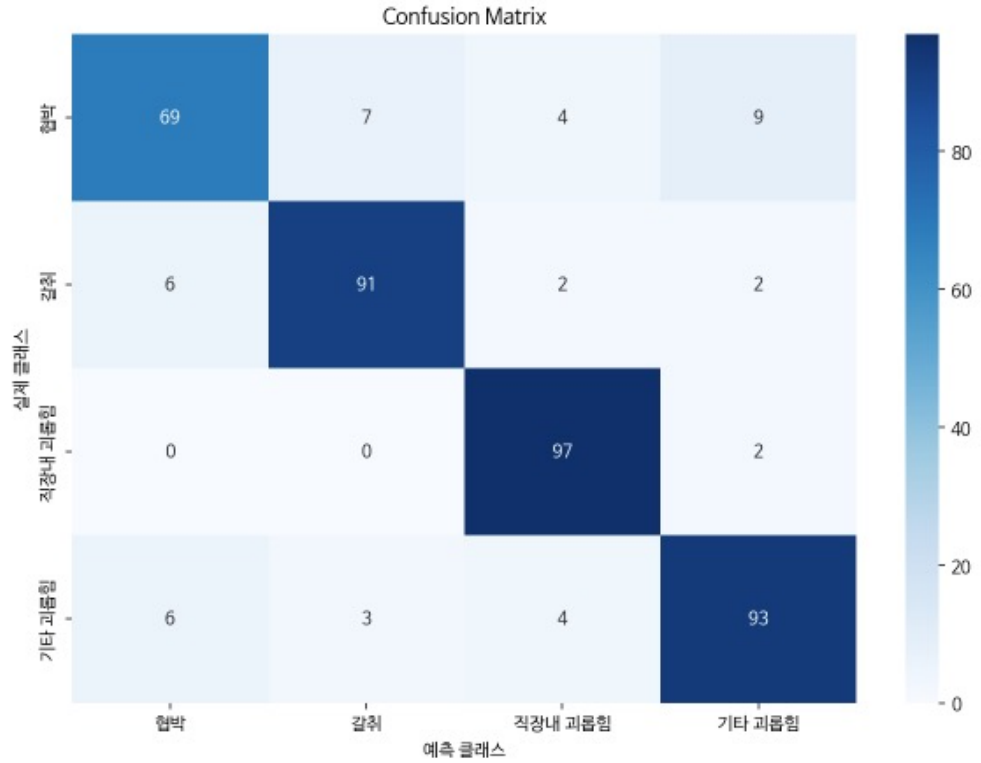
혼동행렬을 통하여 비교한 결과 단순 공백반 토큰화를 통한 모델링보다,
TF_IDF를 이용하여 진행한 학습 결과가 더 좋게 나오는 것을 확인 할 수 있음



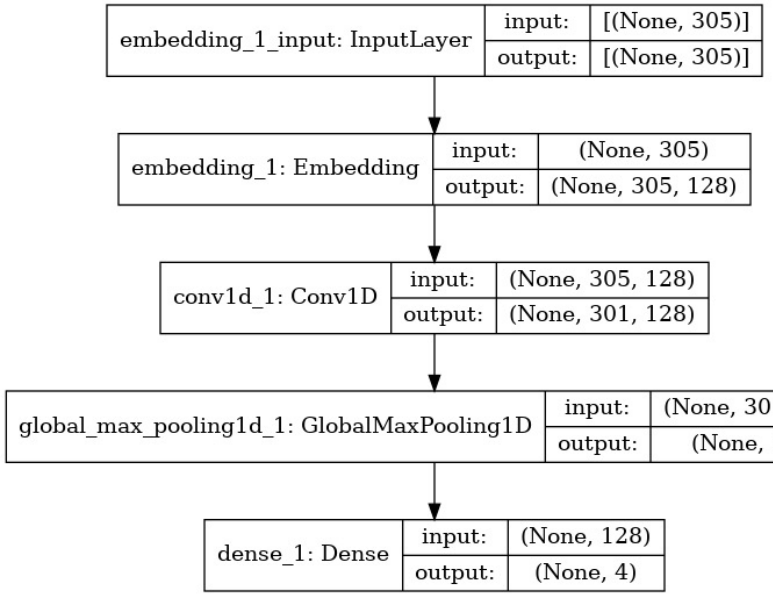
토큰화만 한 경우(좌) & TF-IDF를 이용한 경우(우)

Modelling - CNN

CNN의 경우 1D CNN을 사용하였으며 학습 결과
다른 Class 대비 협박 Class가 성능이 떨어지는 것을 확인되며, 리더보드의 경우 0.865의 성능 기록

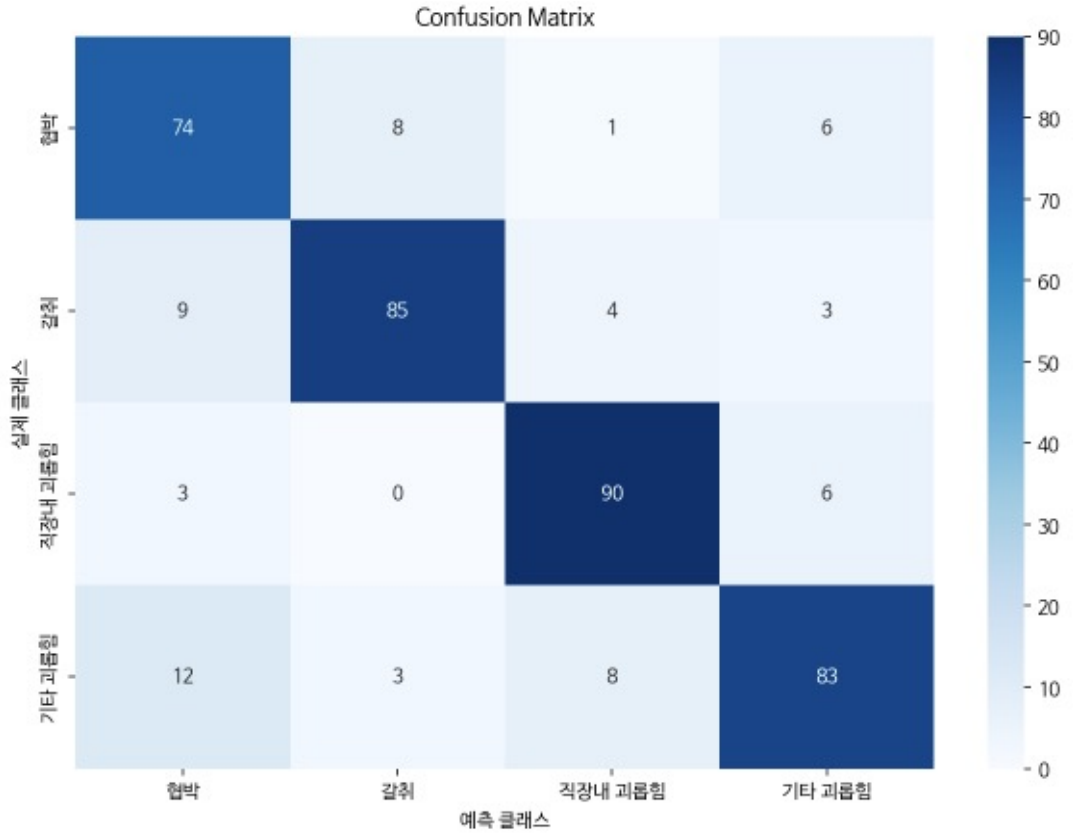


Accuracy: 0.8405
Macro F1-score: 0.8398
Weighted F1-score: 0.8407



Modelling - LSTM

LSTM의 경우 CNN보다 '기타 괴롭힘' Class의 예측이 약했으며,
다른 Class 대비 '협박' Class가 성능이 떨어지는 것을 확인되고, 리더보드의 경우 0.847의 성능 기록



Accuracy: 0.8861
Macro F1-score: 0.8830
Weighted F1-score: 0.8848

embedding_input: InputLayer	input:	[(None, None)]
	output:	[(None, None)]

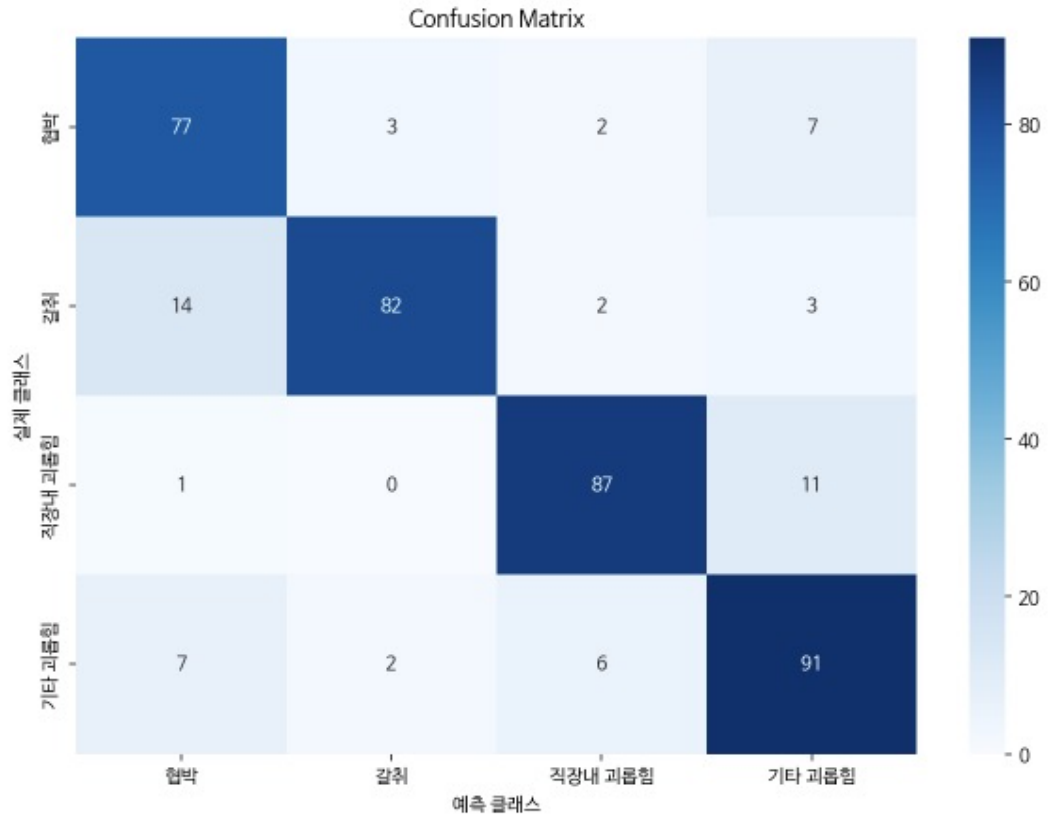
embedding: Embedding	input:	(None, None)
	output:	(None, None, 128)

lstm: LSTM	input:	(None, None, 128)
	output:	(None, 128)

dense: Dense	input:	(None, 128)
	output:	(None, 4)

Modelling - Bi LSTM

Bi LSTM의 경우 LSTM이나 CNN 보다 '협박' Class의 예측이 높았으며,
여전히 다른 Class 대비 '협박' Class가 성능이 떨어지는 것을 확인됨. 리더보드의 경우 0.852의 성능 기록



Accuracy: 0.8532
Macro F1-score: 0.8535
Weighted F1-score: 0.8542

embedding_input: InputLayer	input:	[(None, None)]
	output:	[(None, None)]

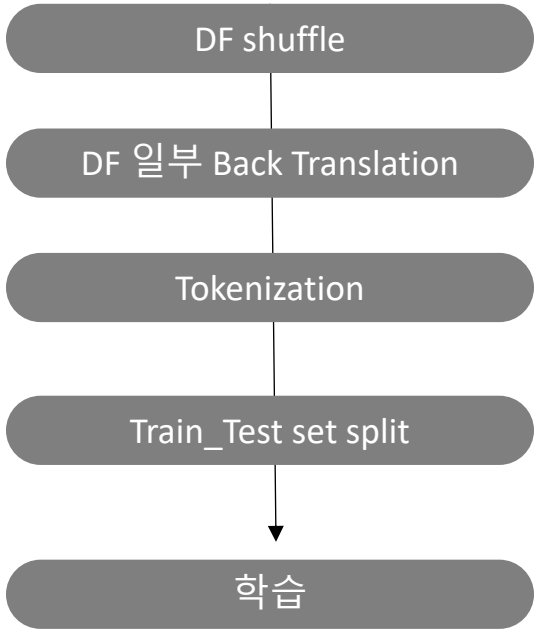
embedding: Embedding	input:	(None, None)
	output:	(None, None, 128)

lstm: LSTM	input:	(None, None, 128)
	output:	(None, 128)

dense: Dense	input:	(None, 128)
	output:	(None, 4)

Modelling - Back Translation

Epoch 마다 Back Translation을 진행하여 Augmentation 진행하여 학습을 진행하였지만
Back Translation 데이터의 품질저하로 인하여 기존 모델 대비 1/3정도 수준의 auc를 보여주었음

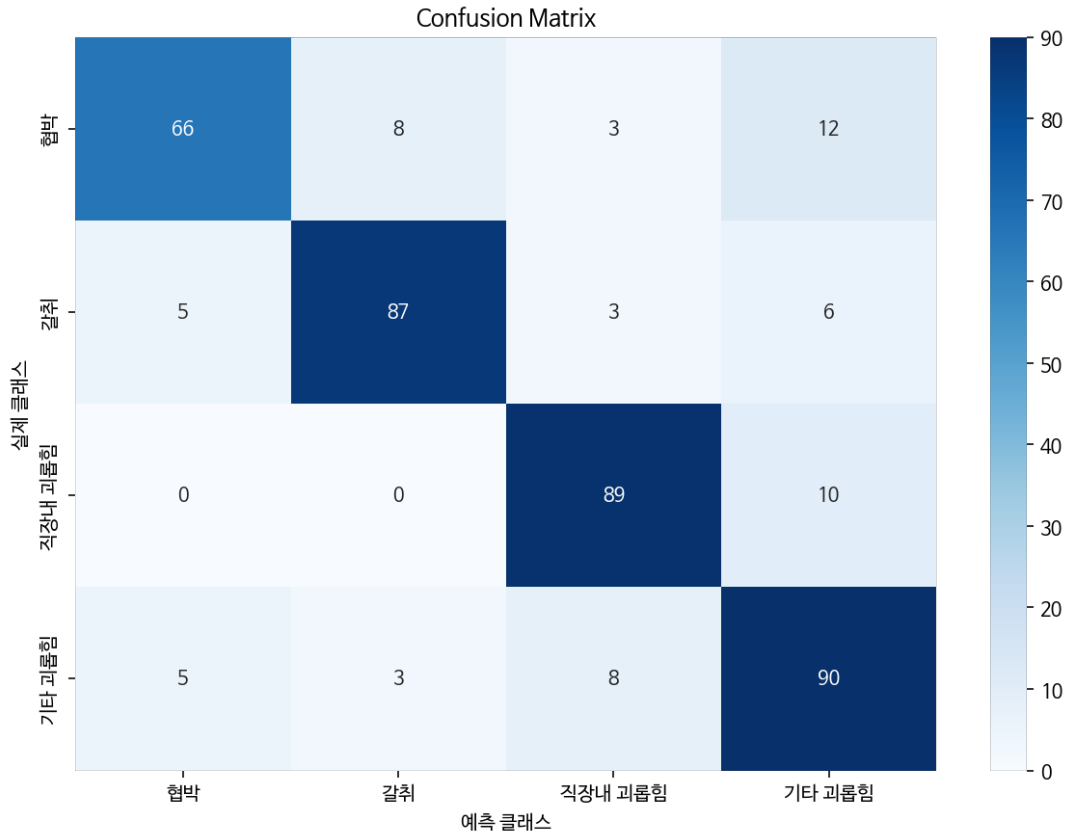


[1epoch task]

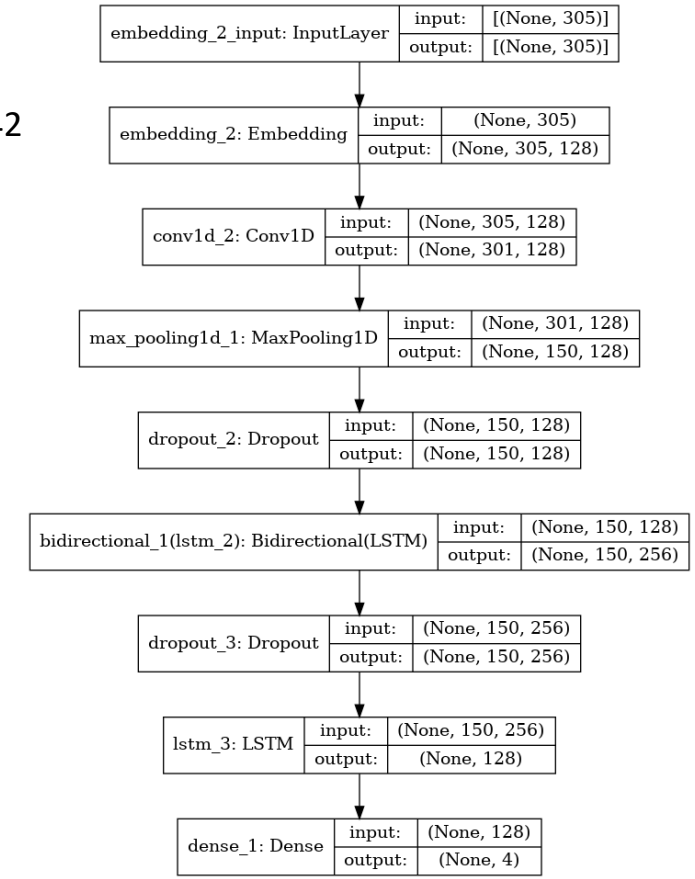
```
Epoch 00001: val_acc did not improve from 0.29620
100%|██████████| 3950/3950 [00:03<00:00, 1286.54it/s]
28/28 [==============================] - 1s 28ms/step - loss: 1.3937 - acc: 0.2608 - val_loss: 1.4065 - val_acc: 0.
2253
Epoch 00001: val_acc did not improve from 0.29620
100%|██████████| 3950/3950 [00:03<00:00, 1043.10it/s]
28/28 [==============================] - 1s 28ms/step - loss: 1.3919 - acc: 0.2577 - val_loss: 1.3958 - val_acc: 0.
2380
Epoch 00001: val_acc did not improve from 0.29620
100%|██████████| 3950/3950 [00:03<00:00, 1269.87it/s]
28/28 [==============================] - 1s 27ms/step - loss: 1.3929 - acc: 0.2568 - val_loss: 1.3883 - val_acc: 0.
2633
Epoch 00001: val_acc did not improve from 0.29620
100%|██████████| 3950/3950 [00:03<00:00, 1282.13it/s]
28/28 [==============================] - 1s 27ms/step - loss: 1.3892 - acc: 0.2686 - val_loss: 1.3860 - val_acc: 0.
2911
Epoch 00001: val_acc did not improve from 0.29620
100%|██████████| 3950/3950 [00:03<00:00, 1291.82it/s]
28/28 [==============================] - 1s 28ms/step - loss: 1.3885 - acc: 0.2689 - val_loss: 1.3897 - val_acc: 0.
2354
Epoch 00001: val_acc did not improve from 0.29620

13/13 [==============================] - 0s 10ms/step - loss: 1.3876 - acc: 0.2608

 bi_lstm_bt 테스트 정확도: 0.2608

13/13 [==============================] - 1s 10ms/step - loss: 1.4071 - acc: 0.2456

 lstm_bt 테스트 정확도: 0.2456
```

Modelling - CNN_LSTM_Bi lstm

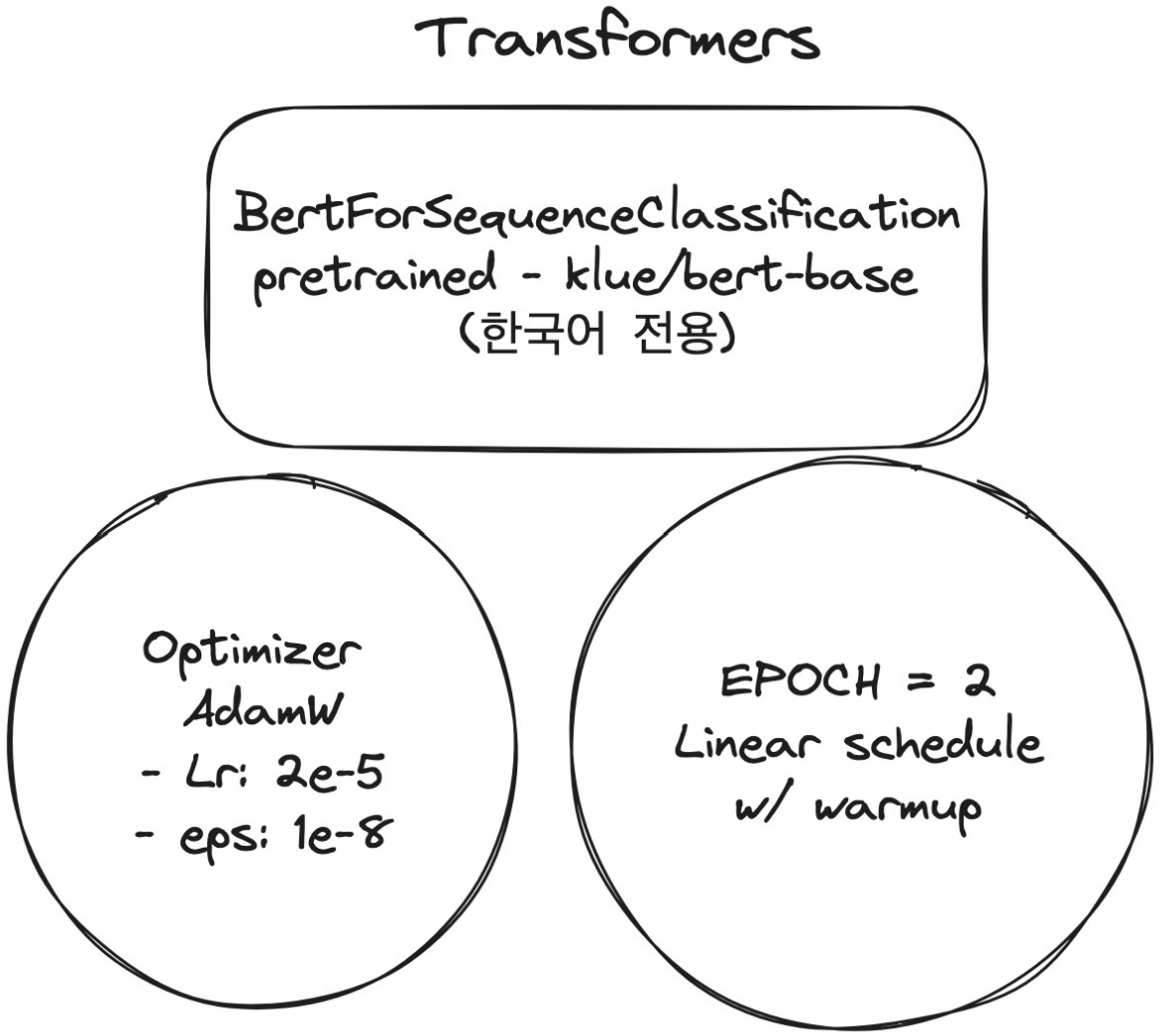
추가적으로 CNN, LSTM, Bi LSTM을 결합한 모델을 학습 진행하였으며
협박 Class의 분류가 부족하긴 하지만 리더보드에서 0.882의 준수한 성적을 보임



Accuracy: 0.8532
Macro F1-score: 0.8535
Weighted F1-score: 0.8542



Modelling - BERT



Accuracy: 0.8747

Macro F1-score: 0.8744

Micro F1-score: 0.8747

Weighted F1-score: 0.8744

Known data (80-20)

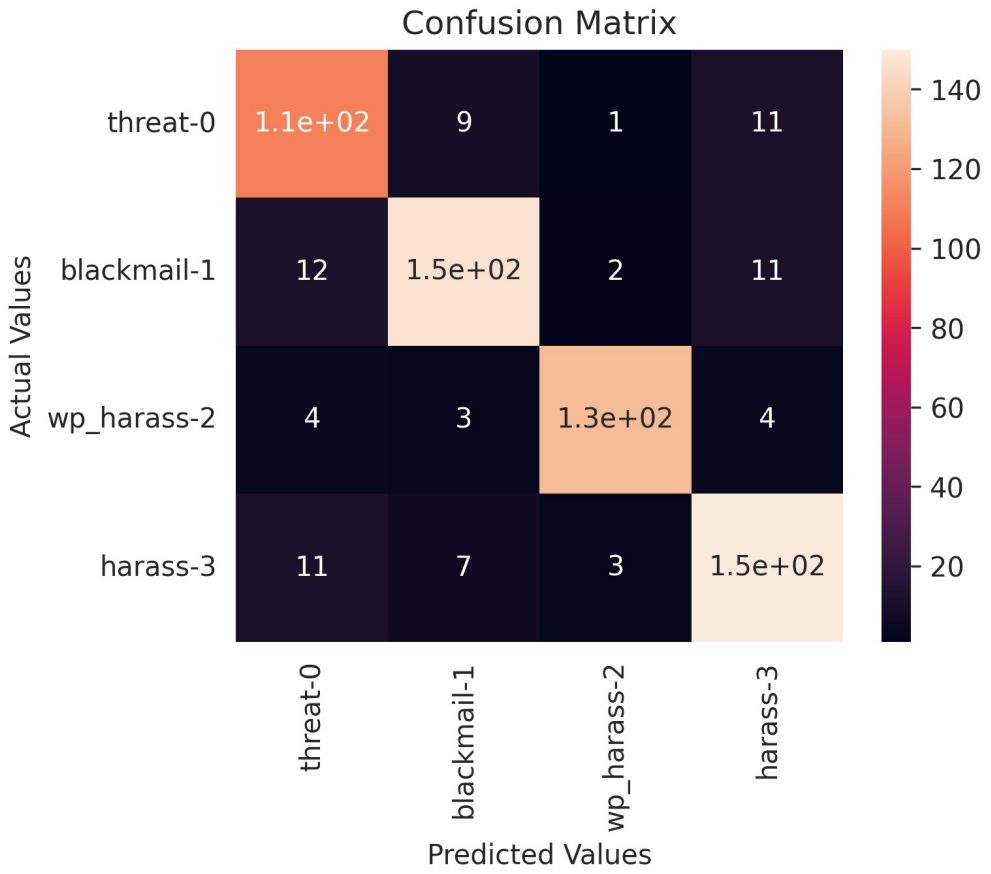
64 % train

16 % valid

20 % test

Modelling - BERT

{협박과 갈취}와 {협박과 기타 괴롭힘}의 경우 모델이 매우 혼동하는 것으로 확인되며,
따라서 협박 클래스의 경우 {갈취, 기타 괴롭힘}, 두 성질을 지니고 있는 것으로 추측 됨
리더보드의 경우 0.87의 score을 기록



Accuracy: 0.8747

Macro F1-score: 0.8744

Micro F1-score: 0.8747

Weighted F1-score: 0.8744

Known data (80-20)

64 % train

16 % valid

20 % test

Conclusion

배운점

전처리 기법들의 구조를 결합하면 각각의 장점을 이용할 수 있지만, 이것이 항상 성능 향상으로 이어지지 않는다는 점을 배울 수 있었다. 하지만, 다양한 모델들을 접할 수 있어 재미있었다.
아쉬운점: 레이어를 순차적으로만 쌓아서 그런지 {협박}에서의 분류 성능이 낮게 나왔다. 일부 레이어를 병합하여 더욱 복잡한 상황에서도 성능 개선을 해 보고 싶다. 그래서인지, {협박} 에 대한 피쳐들을 뽑고 싶어서 xAI에 손을 대었지만, 시간관계상 하기 힘들었다.

느낀점

모델이 {협박}을 제일 헷갈려하는 분류인걸 보면, 협박적인 요소가 우리 사회의 영원한 악플임을 엿볼 수 있었다. 협동 프로젝트이지만 각자 해보고 싶은 부분을 진행을 해서 협업이 조금 안 되었다. 그 과정에서 자기만의 스타일로 코드를 구성하다 보니 코드를 이해하는 부분에도 시간 소모가 꽤 있었다. 이번 기회로 배운 내용을 앞으로 있을 아이펠 톤에 잘 적용해 나가야 겠다는 생각이 들었다.

Q&A