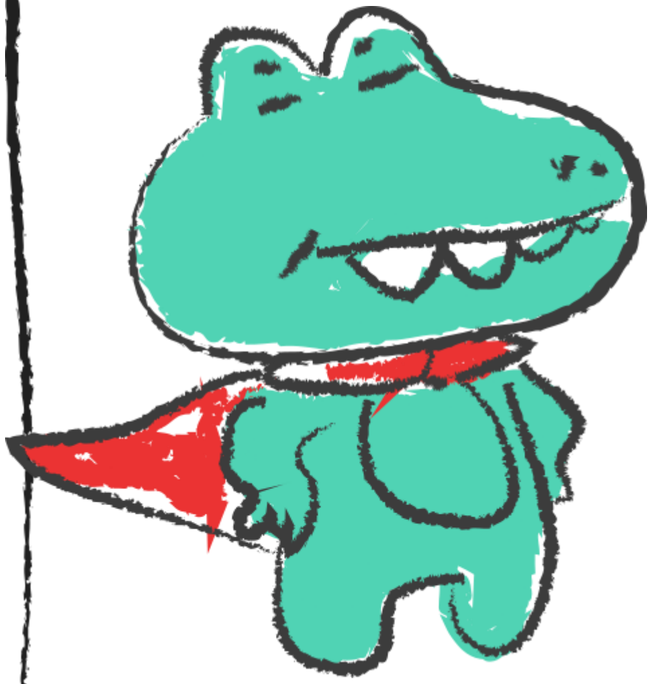
소설 작가 분류 AI 프로젝트



B조

팀원

곽희원

발표자

전처리 및 시각화

Glove

Bert

김홍범

전처리 및 시각화

GRU

CNN개선

선우미

전처리

BI-LSTM

BI-LSTM개선

이원권

전처리 및 시각화

Word2vec

CNN

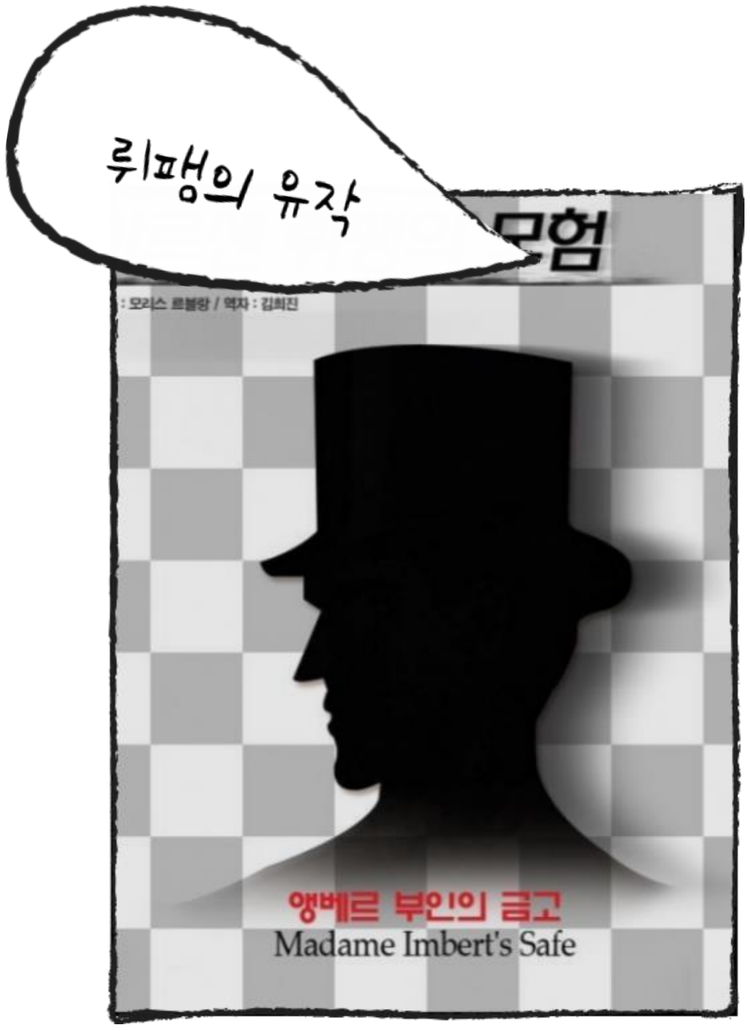
정혜원

전처리 및 시각화

LSTM

LSTM개선

warm up



사라진 유작 진위를 밝힐 수 있을까?

출처-유퀴즈온더튜브

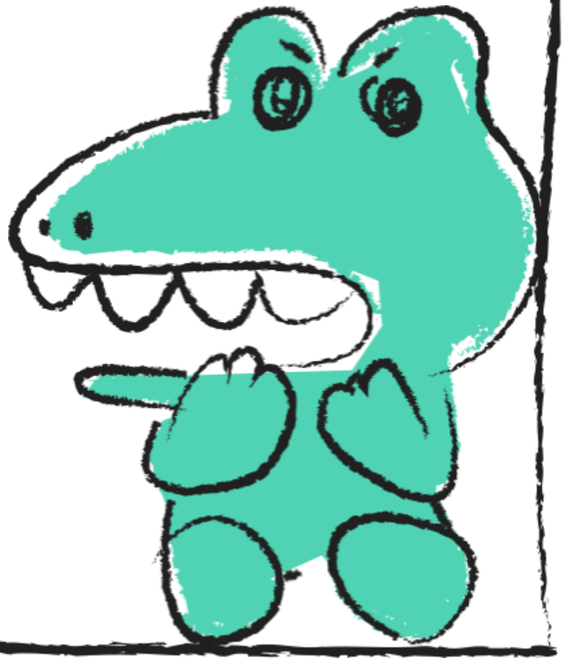
(https://youtu.be/YnP-rHeZEsM)

발표주제

Accuracy상승시키기

성능향상에 가장 좋은

최적의 모델 찾기



목차

1. 작업일지
2. 핵심키워드 소개
3. EDA - 전처리
4. 모델 설명 및 결과값
5. 피드백

작업일지

타임라인

https://www.notion.so/4-24-4-29-8dd83f8af00a4d7db71faaee88008c19

작업일지

팀 작업일지

개인작업일지

소설 작가 분류 작업일지 (4/24 ~ 4/29)

날짜 별 작업 일지

▼ 21-04-24(토)
- 개인 별 DACON 자연어 튜토리얼 학습
- 개인 별 멋쟁이사자처럼 2기 Notion 코드 학습

▼ 21-04-25(일)
- 자연어 처리의 전처리, Vectorization, Embedding 기법을 각각 조사해보고 발표
- 다음주 주간 작업 계획 작성 및 역할 분담
- 개인 별 구축할 모델 선정 및 학습

▼ 21-04-26(월)
- 임베딩 조와 전처리 조로 나누어 각 조 별로 임베딩과 전처리 구현 및 시각화 아이디어
- 임베딩 조는 DACON 파이프라인을 기반으로 Word2Vec과 Glove 임베딩 구현
- 전처리 조는 DACON 기본 제공 및 nltk 라이브러리 기반 전처리와 이전 기수 전처리 구현

▼ 21-04-27(화)
- CNN, LSTM, Bi-LSTM, GRU, BERT의 각 모델 구현
- Word2Vec과 Glove에 맞춰 모델 테스트

▼ 21-04-28(수)
- 모델 학습률 등 하이퍼 파라미터 조정
- 최선의 모델 + 임베딩 기법과 최적화 기법의 조합을 찾기 위해 종류 별로 테스트
- Stopword와 문장 부호를 포함하여 전처리 방법을 새로 정의
- 여러 가지 모델 확인 결과, Word2Vec + Bi-LSTM 모델에 learning rate = 0.001, optimizer =

소설 작가 분류 작업일지 (4/24 ~ 4/29)

이원권

▼ 21-04-24(토)
- **[NLP 언제까지 미룰래? 일단 들어와!!]** 학습 완료

  자연어처리, NLP 전처리, Vectorization, Word Embedding, Modeling 학습

  [NLP 언제까지 미룰래? 일단 들어와!!] #1.자연어 처리란?
  월간 데이콘 9 | 소설 문체 | NLP | Logloss
  https://dacon.io/competitions/official/235670/codeshare/1801...

- NLP

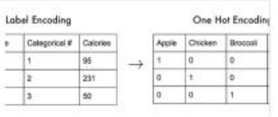
▼ 21-04-25(일)
- **[코드] 데이콘 기초 베이스라인** 구현 완료

  [코드] 데이콘 기초 베이스라인
  월간 데이콘 9 | 소설 문체 | NLP | Logloss
  https://dacon.io/competitions/official/235670/codeshare/1738...

- Word2Vec, Glove, FastText 등 Embedding 기법 조사 및 발표
- 임베딩 기법 예시를 참조하여 각 임베딩 별 code snippet 작성

  임베딩 기법(Embedding)
  본 글은 자연어 처리에서 주로 사용하는 임베딩 기법들에 대해 정리해놓은 자료입니다. One-hot Encoding, TF-IDF, LSA, Word2Vec, Glove, FastText에
  https://eda-ai-lab.tistory.com/428

- 다음주 월요일부터 목요일까지 작업 일정 계획 및 역할 분담

▼ 21-04-26(월)
- 임베딩 조를 맡아 DACON 베이스라인을 기반으로 하여 Word2Vec 임베딩 구현
- 기존 code snippet을 그대로 프로젝트에 적용하기엔 문제가 있어 수정 및 테스트

https://www.notion.so/4-24-4-29-8dd83f8af00a4d7db71faaee88008c19

핵심 키워드

전처리

cleaning(부호제거)

대,소문자 처리

불용어,토큰화 처리

임베딩

word2vec

glove

LSTM

GRU

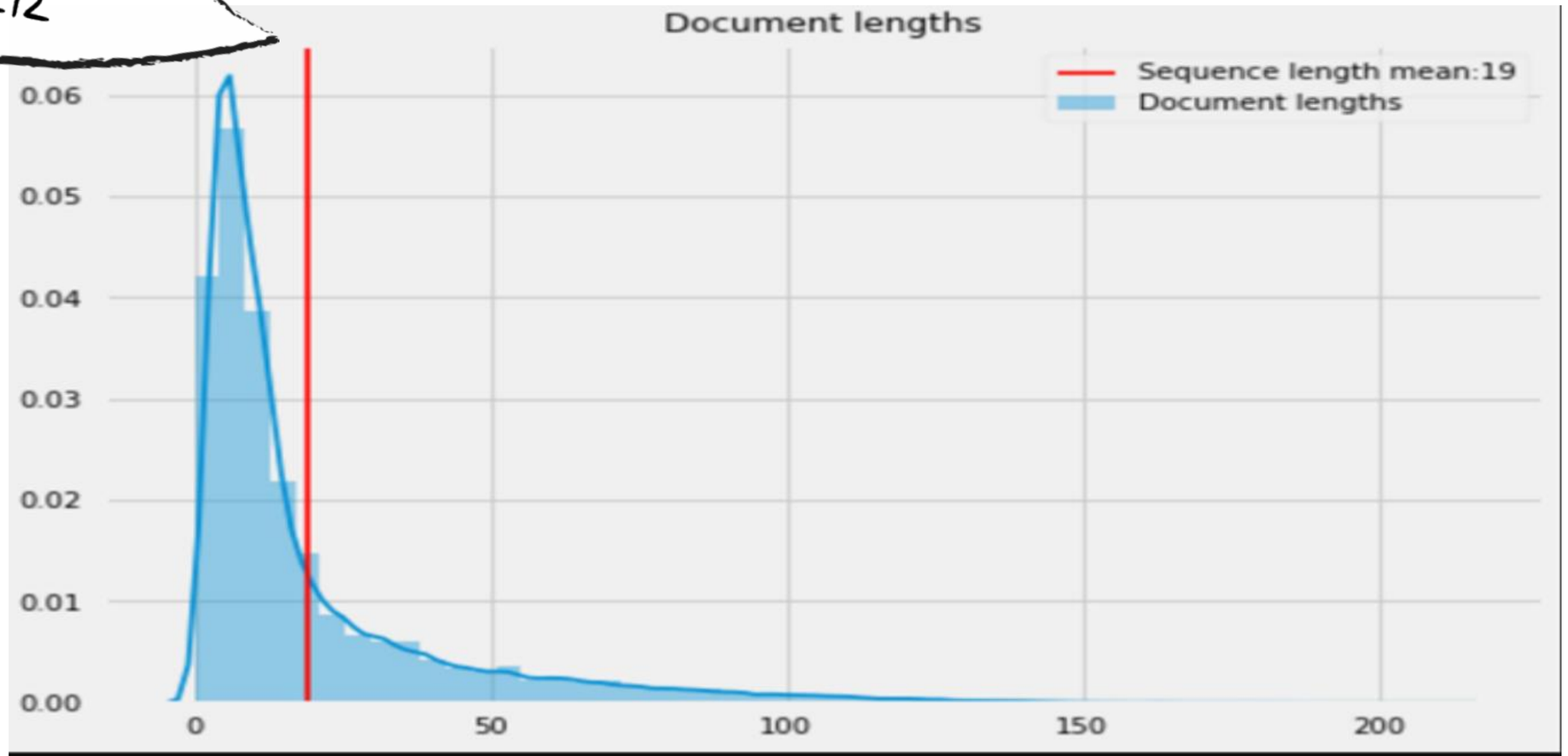
CNN

BI-LSTM

BERT

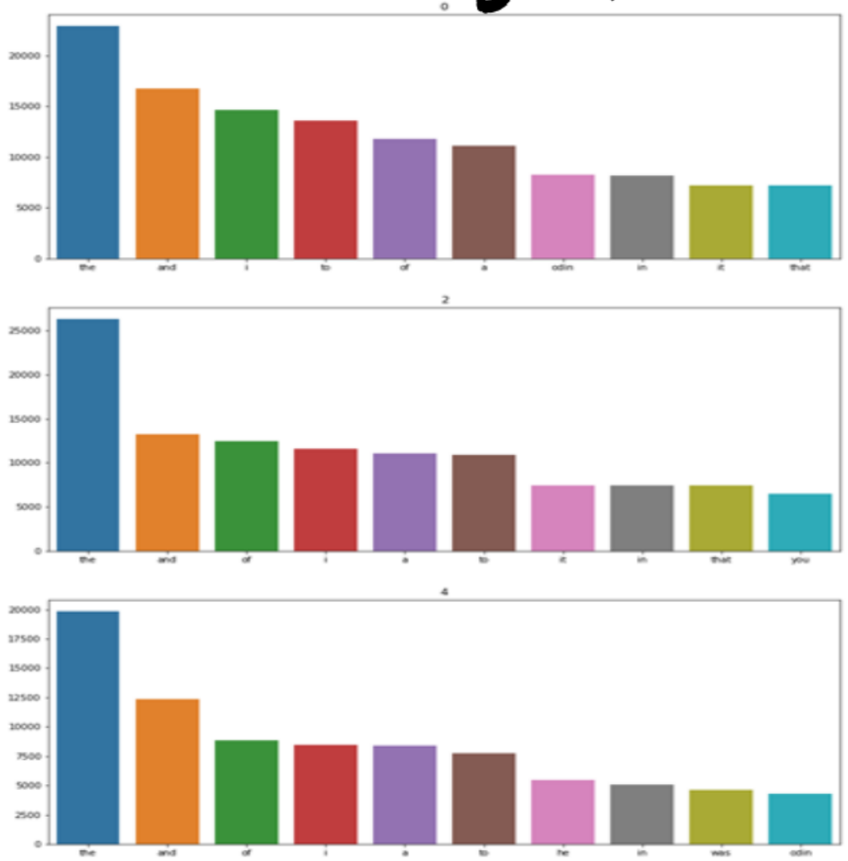
전처리 - 문장길이

max_length = 212



전처리 - 단어 빈도수

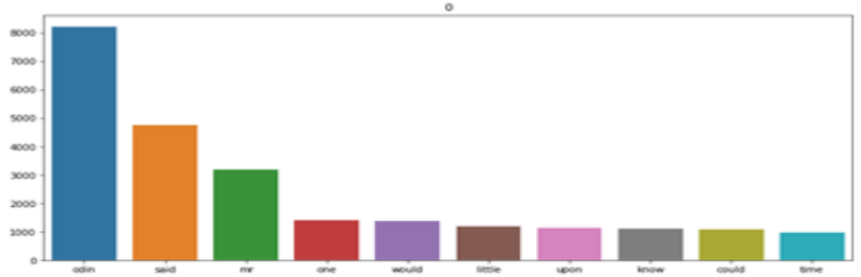
불용어가 많음

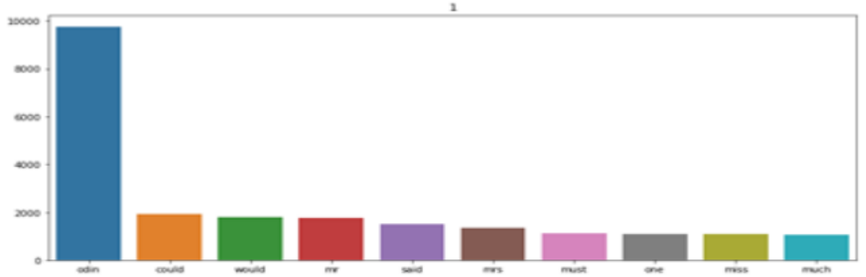


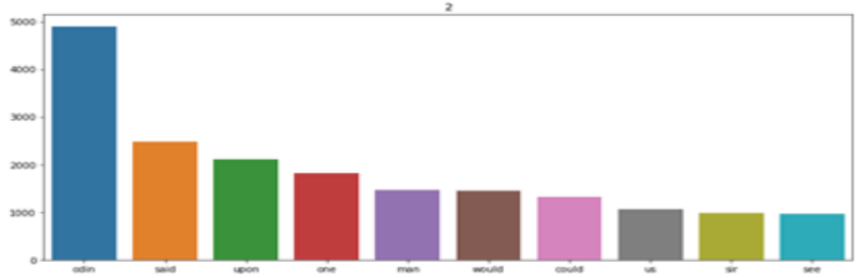
단어 최빈도 top10

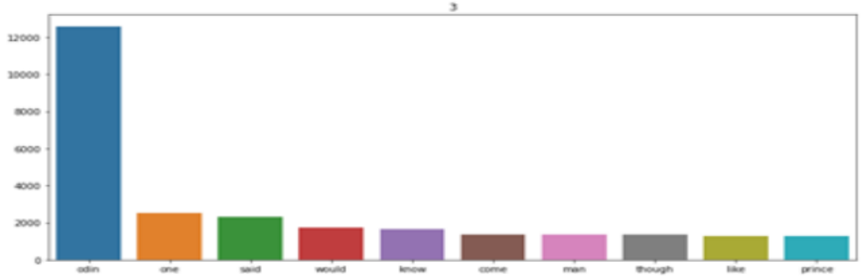
불용어 포함

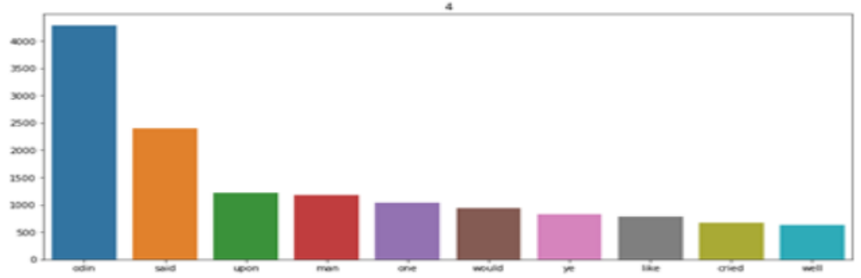
전처리 - 단어 빈도수







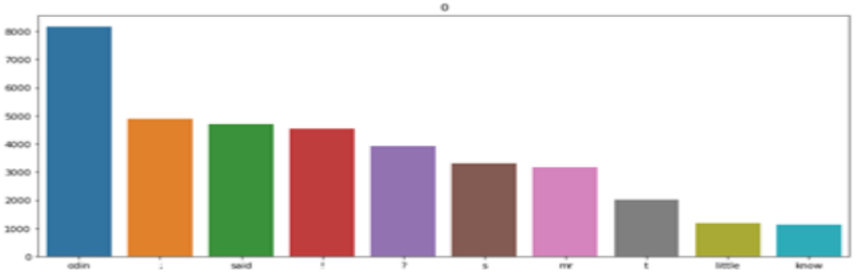


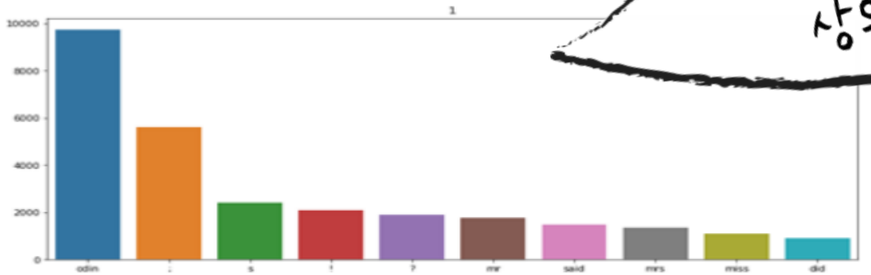


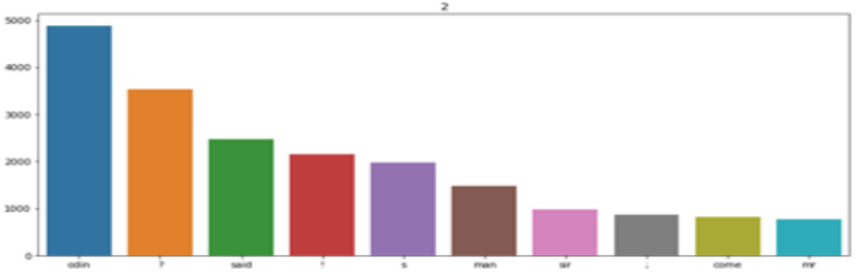
단어 최빈도 top10
불용어 미포함

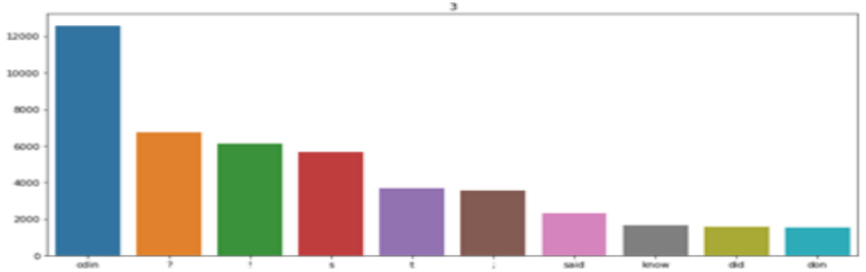
전처리 - 단어 빈도수

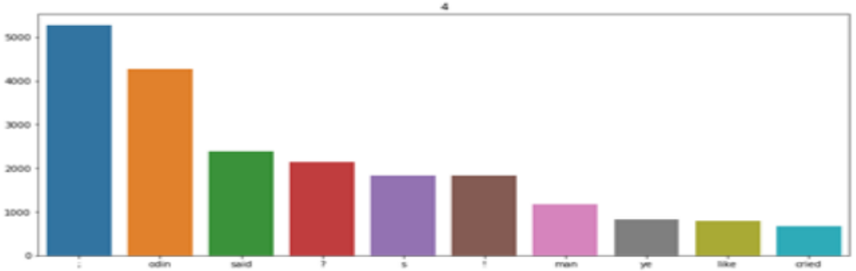
세미콜론이 상위권!







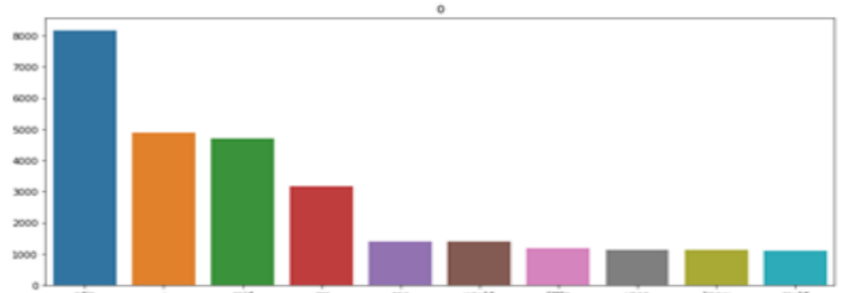


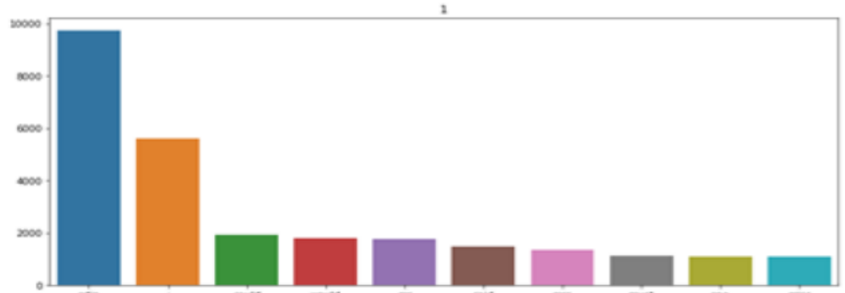


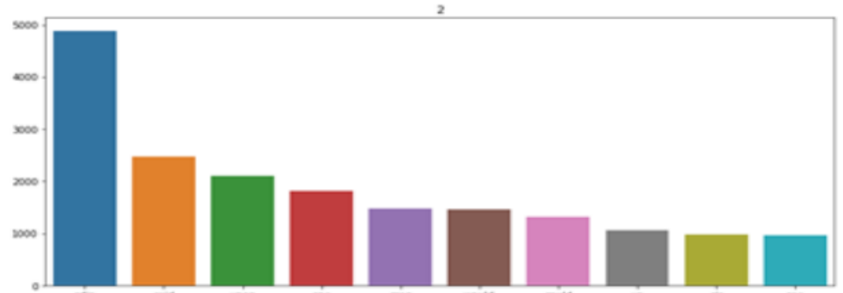
단어 최빈도 top10

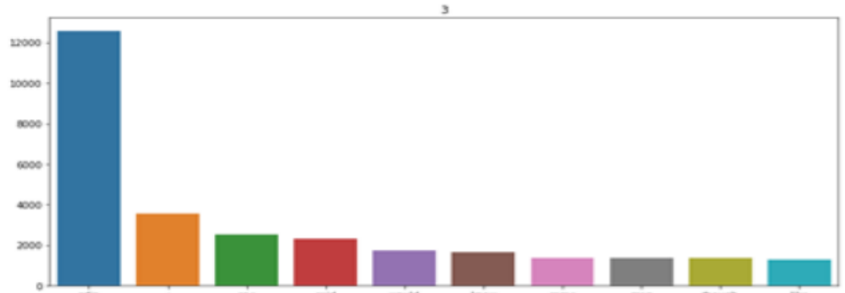
불용어 미포함 / 문장부호 포함

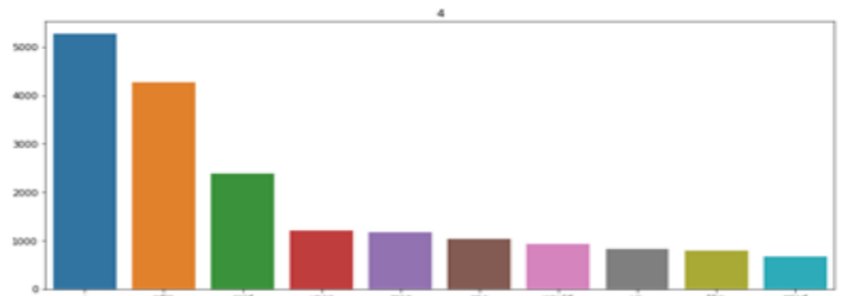
전처리 - 단어 빈도수











단어 최빈도 top10

불용어 미포함 /

콜론,세미콜론 포함

전처리 - 품사

품사의 종류

명사 (noun)

동사 (verb)

형용사 (adjective)

부사 (adverb)

대명사 (pronoun)

한정사 (determiner)

전치사 (preposition)

접속사 (conjunction)

감탄사 (interiection)

주요 품사

그 외 품사

전처리 - 품사

	index	text	author	JJ	RB
0	0	he was almost choking. there was so much, so m...	3	much strange pole evident	almost so so much fixedly
1	1	“your sister asked for it, i suppose?”	2		suppose
2	2	she was engaged one day as she walked, in per...	1	last	not instead again away immediately
3	3	the captain was in the porch, keeping himself ...	4	treacherous dr. north careful	carefully lively
4	4	“have mercy, gentlemen!” odin flung up his han...	3	“ i ve torn	anyway here
		...	...	...	...
		l. “i h...	2		hardly
		n us ...	4		
54876	54876	"your sincere well-wisher, friend, and sister...	1	sincere well-wisher	
54877	54877	“then you wanted me to lend you money?”	3		then
54878	54878	it certainly had not occurred to me before, bu...	0	i	certainly not

POS Tag List in NLTK
형용사 / 부사 추출

전처리 - 품사 (형용사)

Wordcloud by author [0]

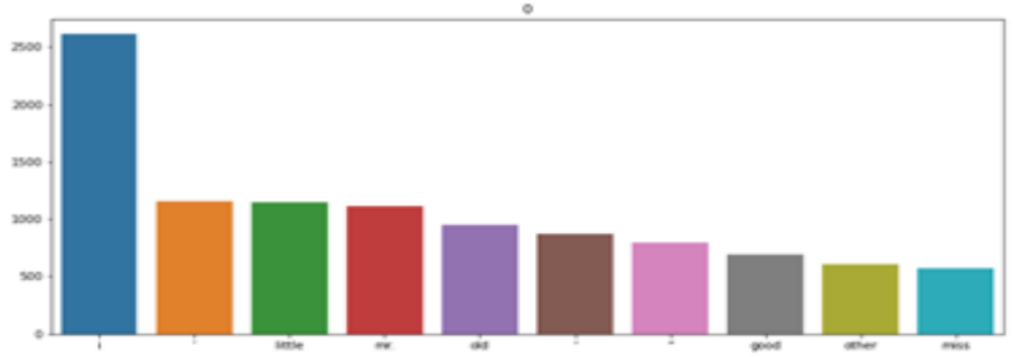
Wordcloud by author [1]

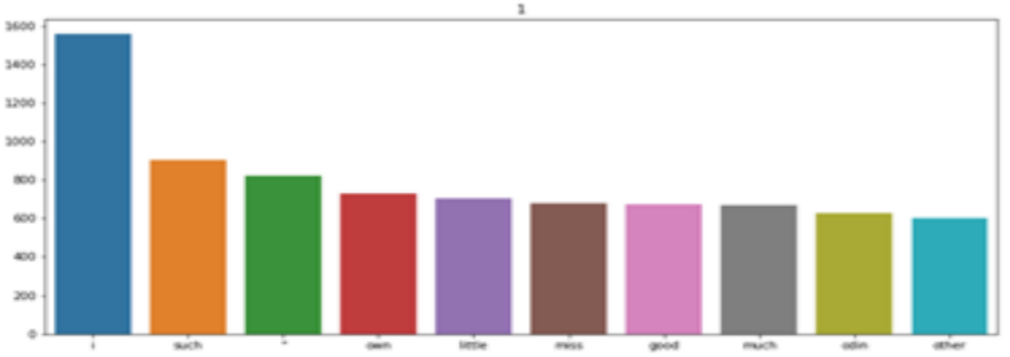
Wordcloud by author [2]

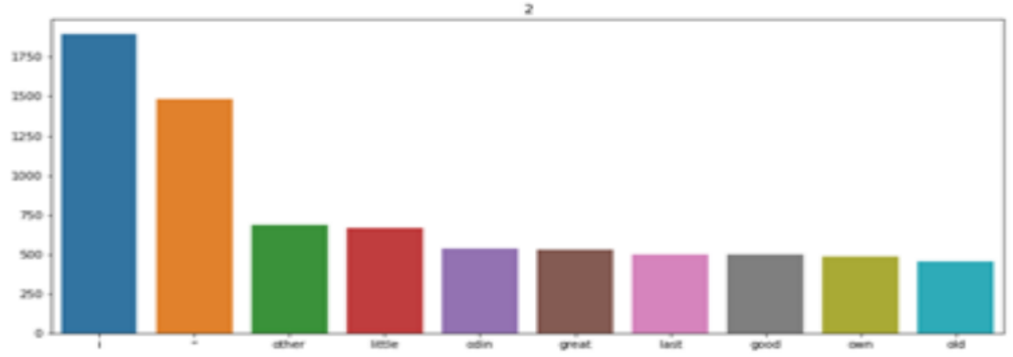
Wordcloud by author [3]

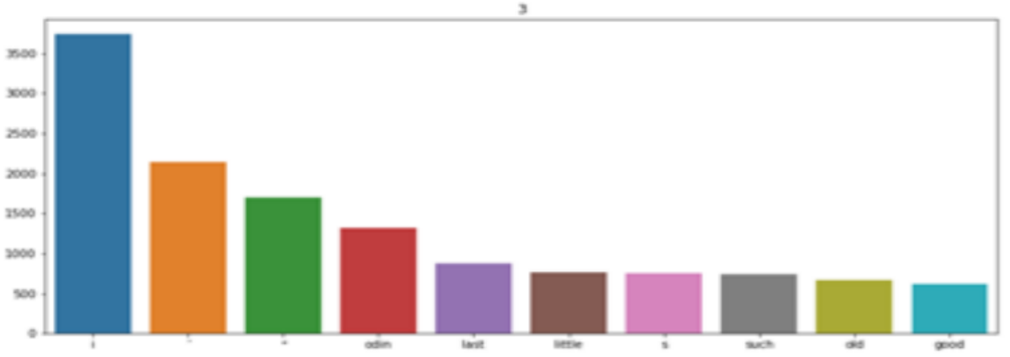
Wordcloud by author [4]

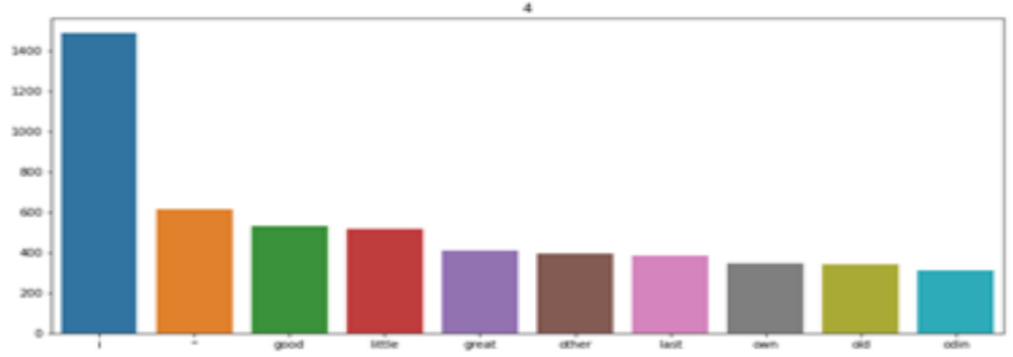
전처리 - 품사 (형용사)









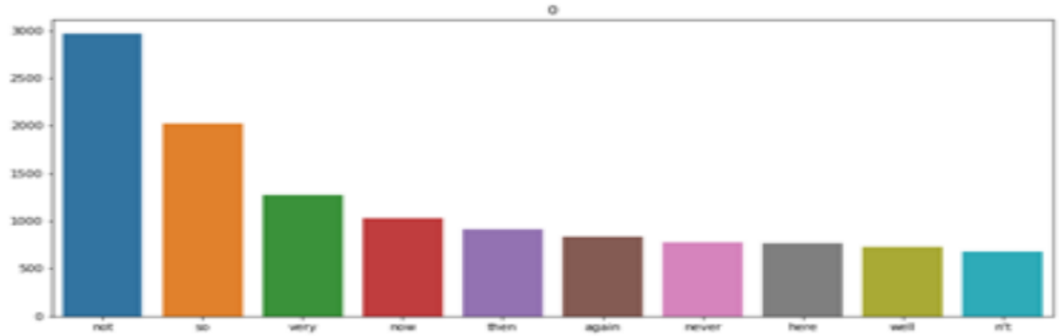


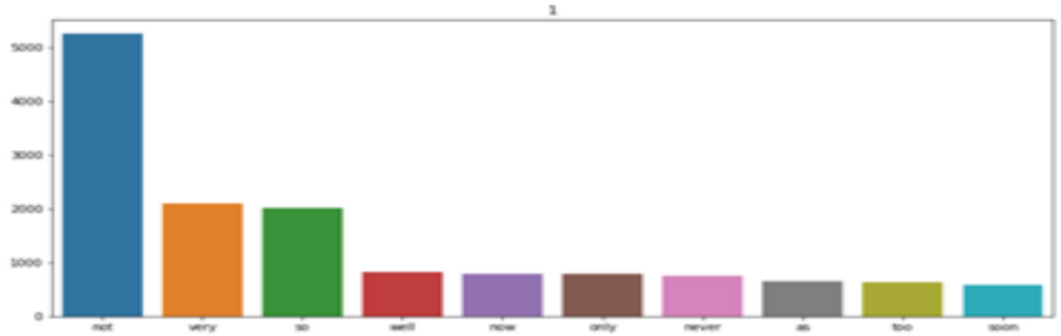
작가	0	1	2	3	4
1등	I	I	I	I	I
2등	‘	such	“	‘	“
3등	little	“	other	“	good

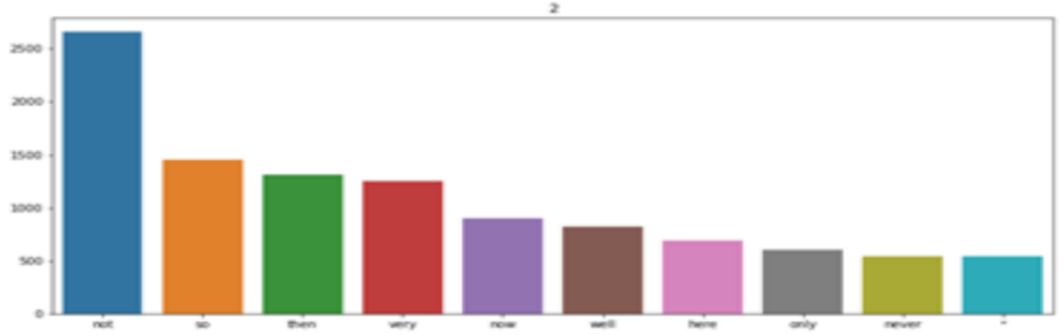
전처리 – 품사 (부사)

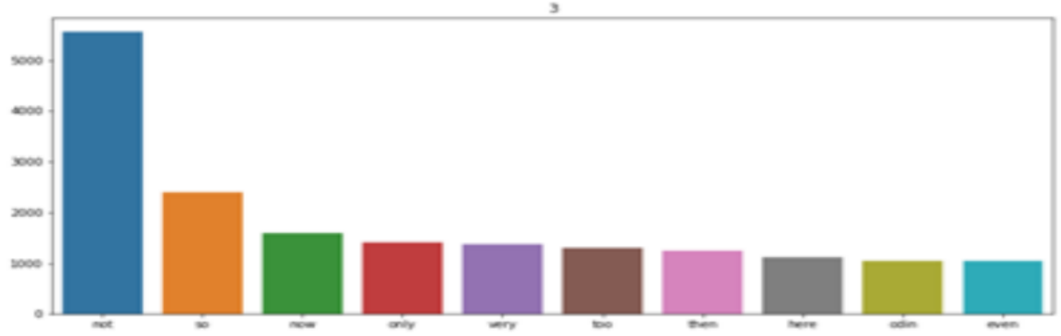

Wordcloud by author [0]
well'
never'
still'
quite'
odin'
n't
back'
away'
ever'
always'
Wordcloud by author [1]
well'
never'
odin'
ever'
however'
quite'
soon'
much'
always'
still'
Wordcloud by author [2]
well'
never'
back'
ever'
even'
however'
still'
odin'
away'
yet'
Wordcloud by author [3]
odin'
even'
suddenly'
never'
still'
well'
Wordcloud by author [4]
well'
still'
never'
even'
back'
perhaps'
indeed'
odin'
yet'
already'

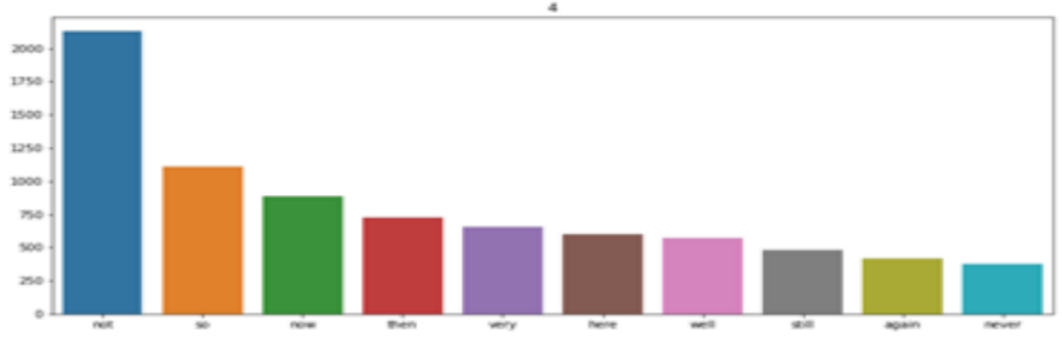
전처리 - 품사 (부사)







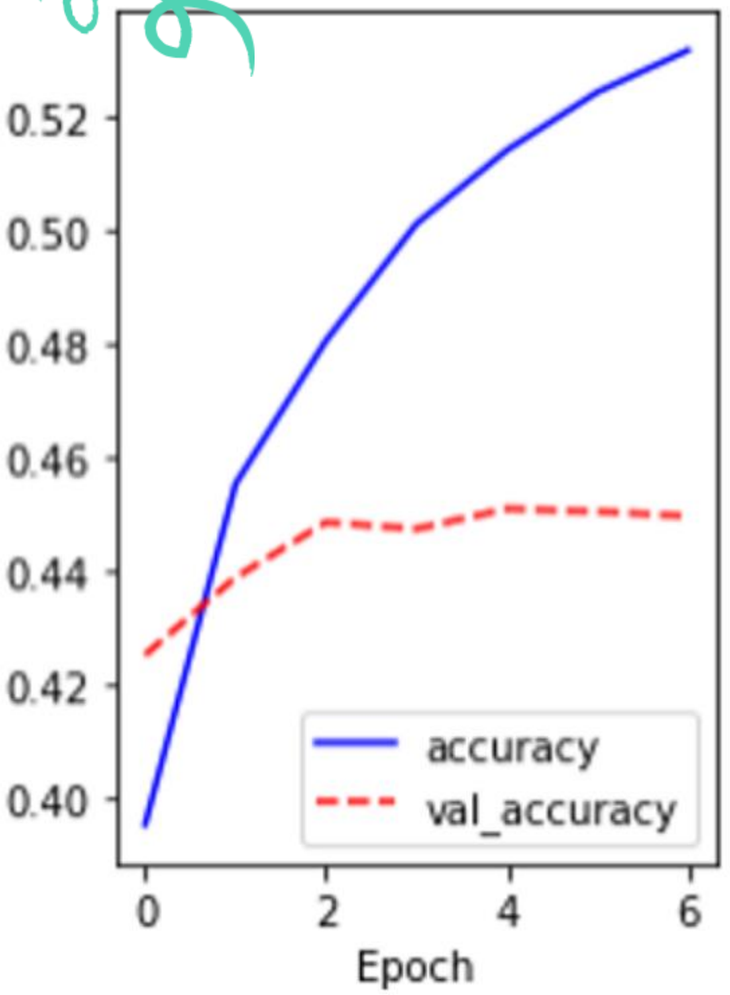




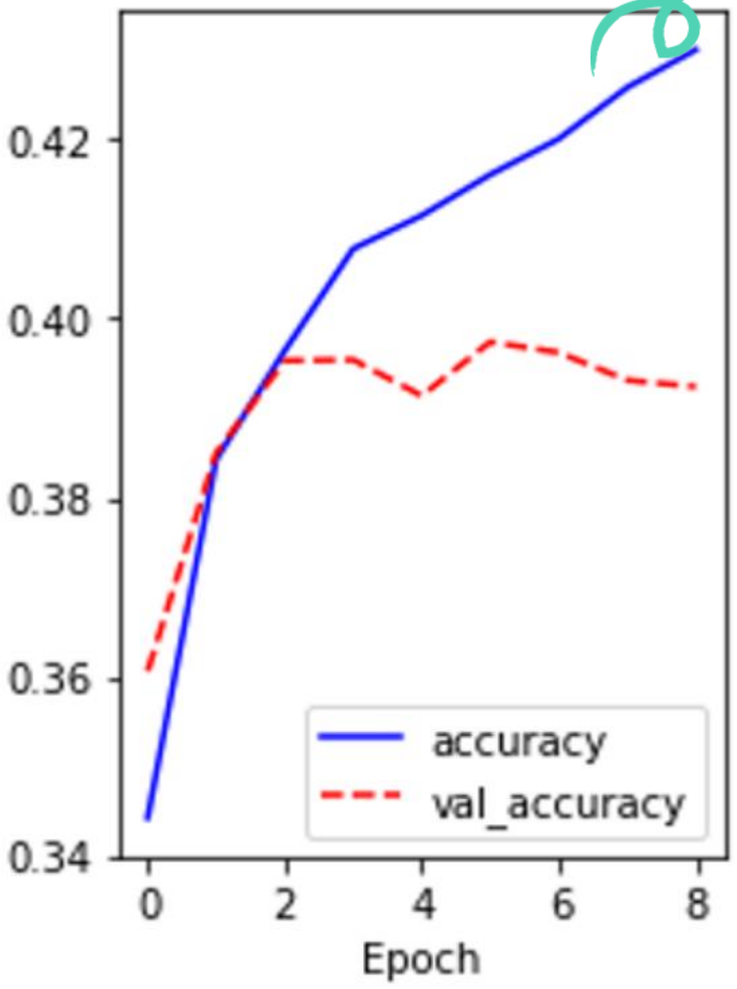
작가	0	1	2	3	4
1등	not	not	not	not	not
2등	so	very	so	so	so
3등	very	so	then	now	now

전처리 - 품사 (형용사/부사 정확도)

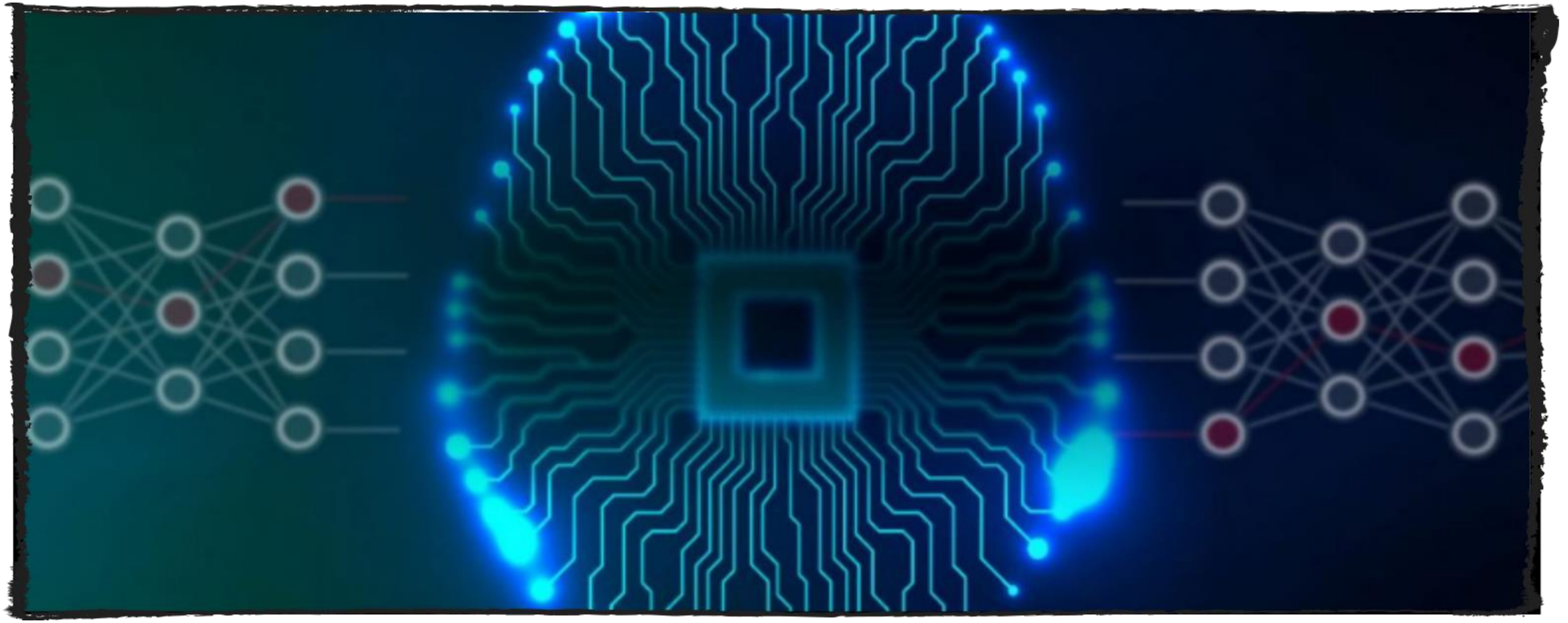
형용사



부사



모델링



모델링 - Word2Vec + LSTM

```
Epoch 1/20
440/440 [==============================] - 305s 680ms/step - loss: 1.5766 - accuracy: 0.2685 - val_loss: 1.5725 - val_accuracy: 0.2680
Epoch 2/20
440/440 [==============================] - 312s 708ms/step - loss: 1.5704 - accuracy: 0.2766 - val_loss: 1.5717 - val_accuracy: 0.2680
Epoch 3/20
440/440 [==============================] - 312s 709ms/step - loss: 1.5698 - accuracy: 0.2786 - val_loss: 1.5728 - val_accuracy: 0.2680
Epoch 4/20
440/440 [==============================] - 313s 711ms/step - loss: 1.5703 - accuracy: 0.2769 - val_loss: 1.5711 - val_accuracy: 0.2680
Epoch 5/20
440/440 [==============================] - 313s 712ms/step - loss: 1.5736 - accuracy: 0.2763 - val_loss: 1.5715 - val_accuracy: 0.2680
Epoch 6/20
440/440 [==============================] - 316s 719ms/step - loss: 1.5695 - accuracy: 0.2800 - val_loss: 1.5713 - val_accuracy: 0.2680
Epoch 7/20
440/440 [==============================] - 313s 712ms/step - loss: 1.5695 - accuracy: 0.2769 - val_loss: 1.5715 - val_accuracy: 0.2680
Epoch 8/20
440/440 [==============================] - 313s 712ms/step - loss: 1.5710 - accuracy: 0.2760 - val_loss: 1.5715 - val_accuracy: 0.2680
Epoch 9/20
440/440 [==============================] - 313s 711ms/step - loss: 1.5717 - accuracy: 0.2759 - val_loss: 1.5714 - val_accuracy: 0.2680
Epoch 10/20
440/440 [==============================] - 311s 708ms/step - loss: 1.5689 - accuracy: 0.2769 - val_loss: 1.5718 - val_accuracy: 0.2680
Epoch 11/20
440/440 [==============================] - 312s 710ms/step - loss: 1.5716 - accuracy: 0.2739 - val_loss: 1.5714 - val_accuracy: 0.2680
Epoch 12/20
440/440 [==============================] - 311s 706ms/step - loss: 1.5695 - accuracy: 0.2782 - val_loss: 1.5714 - val_accuracy: 0.2680
Epoch 13/20
440/440 [==============================] - 310s 704ms/step - loss: 1.5704 - accuracy: 0.2741 - val_loss: 1.5712 - val_accuracy: 0.2680
Epoch 14/20
440/440 [==============================] - 310s 704ms/step - loss: 1.5708 - accuracy: 0.2756 - val_loss: 1.5716 - val_accuracy: 0.2680
Epoch 15/20
440/440 [==============================] - 311s 707ms/step - loss: 1.5692 - accuracy: 0.2796 - val_loss: 1.5714 - val_accuracy: 0.2680
Epoch 16/20
440/440 [==============================] - 310s 704ms/step - loss: 1.5694 - accuracy: 0.2795 - val_loss: 1.5714 - val_accuracy: 0.2680
Epoch 17/20
440/440 [==============================] - 309s 703ms/step - loss: 1.5703 - accuracy: 0.2756 - val_loss: 1.5714 - val_accuracy: 0.2680
Epoch 18/20
440/440 [==============================] - 309s 702ms/step - loss: 1.5687 - accuracy: 0.2796 - val_loss: 1.5717 - val_accuracy: 0.2680
Epoch 19/20
440/440 [==============================] - 308s 700ms/step - loss: 1.5716 - accuracy: 0.2742 - val_loss: 1.5716 - val_accuracy: 0.2680
Epoch 20/20
440/440 [==============================] - 309s 702ms/step - loss: 1.5708 - accuracy: 0.2760 - val_loss: 1.5711 - val_accuracy: 0.2680
```

모델링 - Glove + LSTM

```
Epoch 1/20
440/440 [==============================] - 298s 664ms/step - loss: 1.5738 - accuracy: 0.2776 - val_loss: 1.5719 - val_accuracy: 0.2680
Epoch 2/20
440/440 [==============================] - 313s 712ms/step - loss: 1.5704 - accuracy: 0.2754 - val_loss: 1.5715 - val_accuracy: 0.2680
Epoch 3/20
440/440 [==============================] - 314s 715ms/step - loss: 1.5717 - accuracy: 0.2739 - val_loss: 1.5736 - val_accuracy: 0.2680
Epoch 4/20
440/440 [==============================] - 320s 727ms/step - loss: 1.5725 - accuracy: 0.2725 - val_loss: 1.5714 - val_accuracy: 0.2680
Restoring model weights from the end of the best epoch.
Epoch 00004: early stopping
```

모델링 - Word2Vec + LSTM (긍부정)

```
Epoch 1/20
440/440 [==============================] - 290s 646ms/step - loss: 1.5762 - accuracy: 0.2691 - val_loss: 1.5722 - val_accuracy: 0.2680
Epoch 2/20
440/440 [==============================] - 307s 697ms/step - loss: 1.5709 - accuracy: 0.2778 - val_loss: 1.5712 - val_accuracy: 0.2680
Epoch 3/20
440/440 [==============================] - 306s 696ms/step - loss: 1.5685 - accuracy: 0.2796 - val_loss: 1.5716 - val_accuracy: 0.2680
Epoch 4/20
440/440 [==============================] - 308s 701ms/step - loss: 1.5738 - accuracy: 0.2710 - val_loss: 1.5720 - val_accuracy: 0.2680
Epoch 5/20
440/440 [==============================] - 309s 703ms/step - loss: 1.5704 - accuracy: 0.2783 - val_loss: 1.5713 - val_accuracy: 0.2680
Epoch 6/20
440/440 [==============================] - 309s 702ms/step - loss: 1.5675 - accuracy: 0.2795 - val_loss: 1.5711 - val_accuracy: 0.2680
Epoch 7/20
440/440 [==============================] - 309s 702ms/step - loss: 1.5713 - accuracy: 0.2761 - val_loss: 1.5713 - val_accuracy: 0.2680
Epoch 8/20
440/440 [==============================] - 309s 702ms/step - loss: 1.5707 - accuracy: 0.2736 - val_loss: 1.5722 - val_accuracy: 0.2680
Epoch 9/20
440/440 [==============================] - 310s 705ms/step - loss: 1.5694 - accuracy: 0.2777 - val_loss: 1.5719 - val_accuracy: 0.2680
Epoch 10/20
440/440 [==============================] - 307s 697ms/step - loss: 1.5689 - accuracy: 0.2768 - val_loss: 1.5715 - val_accuracy: 0.2680
Epoch 11/20
440/440 [==============================] - 309s 702ms/step - loss: 1.5724 - accuracy: 0.2727 - val_loss: 1.5716 - val_accuracy: 0.2680
Epoch 12/20
440/440 [==============================] - 309s 703ms/step - loss: 1.5721 - accuracy: 0.2749 - val_loss: 1.5714 - val_accuracy: 0.2680
Epoch 13/20
440/440 [==============================] - 309s 702ms/step - loss: 1.5702 - accuracy: 0.2759 - val_loss: 1.5713 - val_accuracy: 0.2680
Epoch 14/20
440/440 [==============================] - 310s 706ms/step - loss: 1.5706 - accuracy: 0.2748 - val_loss: 1.5716 - val_accuracy: 0.2680
Epoch 15/20
440/440 [==============================] - 308s 700ms/step - loss: 1.5702 - accuracy: 0.2739 - val_loss: 1.5716 - val_accuracy: 0.2680
Epoch 16/20
440/440 [==============================] - 309s 702ms/step - loss: 1.5701 - accuracy: 0.2746 - val_loss: 1.5718 - val_accuracy: 0.2680
Epoch 17/20
440/440 [==============================] - 310s 705ms/step - loss: 1.5664 - accuracy: 0.2802 - val_loss: 1.5713 - val_accuracy: 0.2680
Epoch 18/20
440/440 [==============================] - 310s 704ms/step - loss: 1.5696 - accuracy: 0.2759 - val_loss: 1.5715 - val_accuracy: 0.2680
Epoch 19/20
440/440 [==============================] - 312s 710ms/step - loss: 1.5724 - accuracy: 0.2728 - val_loss: 1.5713 - val_accuracy: 0.2680
Epoch 20/20
440/440 [==============================] - 312s 710ms/step - loss: 1.5680 - accuracy: 0.2776 - val_loss: 1.5713 - val_accuracy: 0.2680
```

모델링 - Glove + LSTM (개선)

```
Epoch 1/20
440/440 [==============================] - 312s 696ms/step - loss: 1.5748 - accuracy: 0.2705 - val_loss: 1.5724 - val_accuracy: 0.2680
Epoch 2/20
440/440 [==============================] - 322s 733ms/step - loss: 1.5716 - accuracy: 0.2754 - val_loss: 1.5712 - val_accuracy: 0.2680
Epoch 3/20
440/440 [==============================] - 324s 737ms/step - loss: 1.5710 - accuracy: 0.2775 - val_loss: 1.5717 - val_accuracy: 0.2680
Epoch 4/20
440/440 [==============================] - 322s 731ms/step - loss: 1.5707 - accuracy: 0.2754 - val_loss: 1.5717 - val_accuracy: 0.2680
Epoch 5/20
440/440 [==============================] - 321s 729ms/step - loss: 1.5717 - accuracy: 0.2767 - val_loss: 1.5713 - val_accuracy: 0.2680
Restoring model weights from the end of the best epoch.
Epoch 00005: early stopping
```

모델링 - Word2Vec + GRU

```
Epoch 1/5
86/86 [==============================] - 154s 2s/step - loss: nan - accuracy: 0.2297 - val_loss:
nan - val_accuracy: 0.2472
Epoch 2/5
86/86 [==============================] - 173s 2s/step - loss: nan - accuracy: 0.2396 - val_loss:
nan - val_accuracy: 0.2472
Epoch 3/5
86/86 [==============================] - 181s 2s/step - loss: nan - accuracy: 0.2427 - val_loss:
nan - val_accuracy: 0.2472
Epoch 4/5
86/86 [==============================] - 180s 2s/step - loss: nan - accuracy: 0.2374 - val_loss:
nan - val_accuracy: 0.2472
Epoch 5/5
86/86 [==============================] - 180s 2s/step - loss: nan - accuracy: 0.2384 - val_loss:
nan - val_accuracy: 0.2472
```

모델링 - Word2Vec + CNN

```
Epoch 1/20
1372/1372 - 202s - loss: 1.2567 - accuracy: 0.4722 - val_loss: 0.9826 - val_accuracy: 0.6252
Epoch 2/20
1372/1372 - 202s - loss: 0.8520 - accuracy: 0.6761 - val_loss: 0.8271 - val_accuracy: 0.6885
Epoch 3/20
1372/1372 - 201s - loss: 0.6363 - accuracy: 0.7680 - val_loss: 0.8090 - val_accuracy: 0.7086
Epoch 4/20
1372/1372 - 205s - loss: 0.4890 - accuracy: 0.8233 - val_loss: 0.8066 - val_accuracy: 0.7126
Epoch 5/20
1372/1372 - 202s - loss: 0.3996 - accuracy: 0.8561 - val_loss: 0.8778 - val_accuracy: 0.7147
Epoch 6/20
1372/1372 - 201s - loss: 0.3318 - accuracy: 0.8813 - val_loss: 0.9464 - val_accuracy: 0.7096
Epoch 7/20
1372/1372 - 201s - loss: 0.2881 - accuracy: 0.8980 - val_loss: 1.1080 - val_accuracy: 0.7150
Restoring model weights from the end of the best epoch.
Epoch 00007: early stopping
```

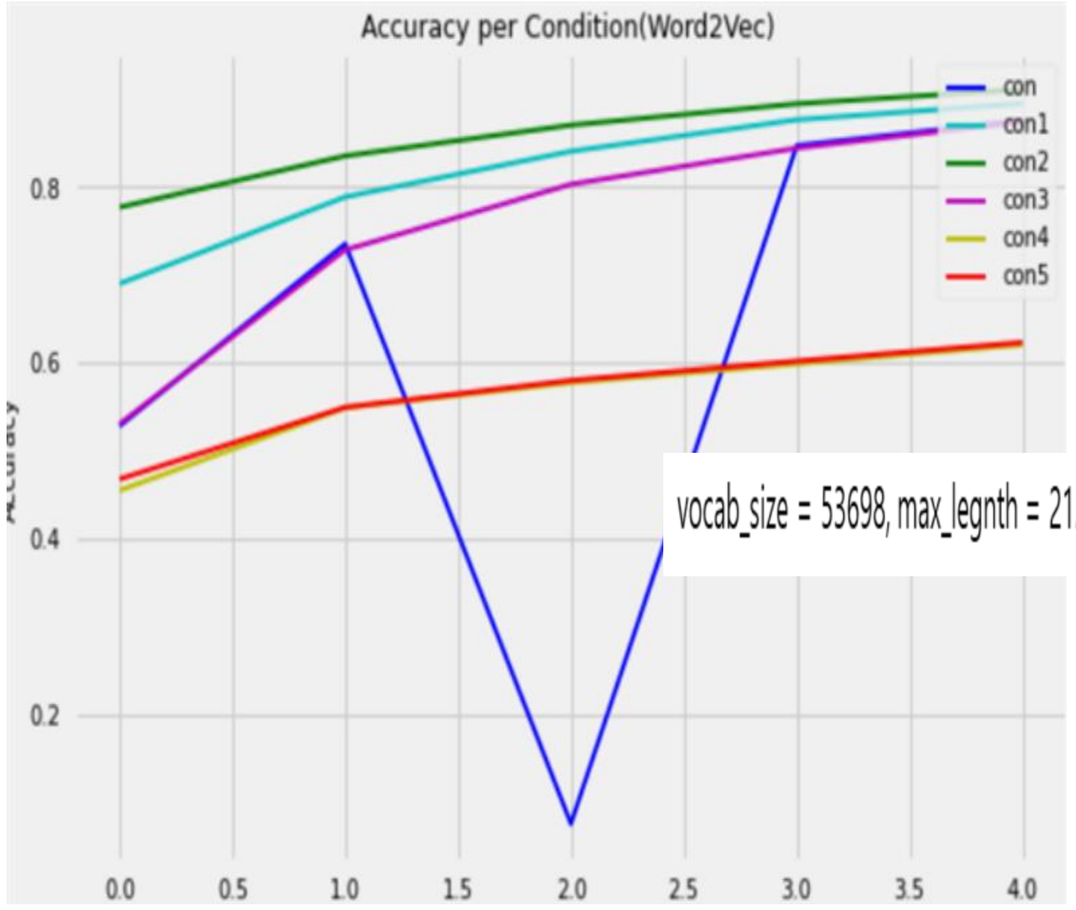
모델링 - Glove + CNN

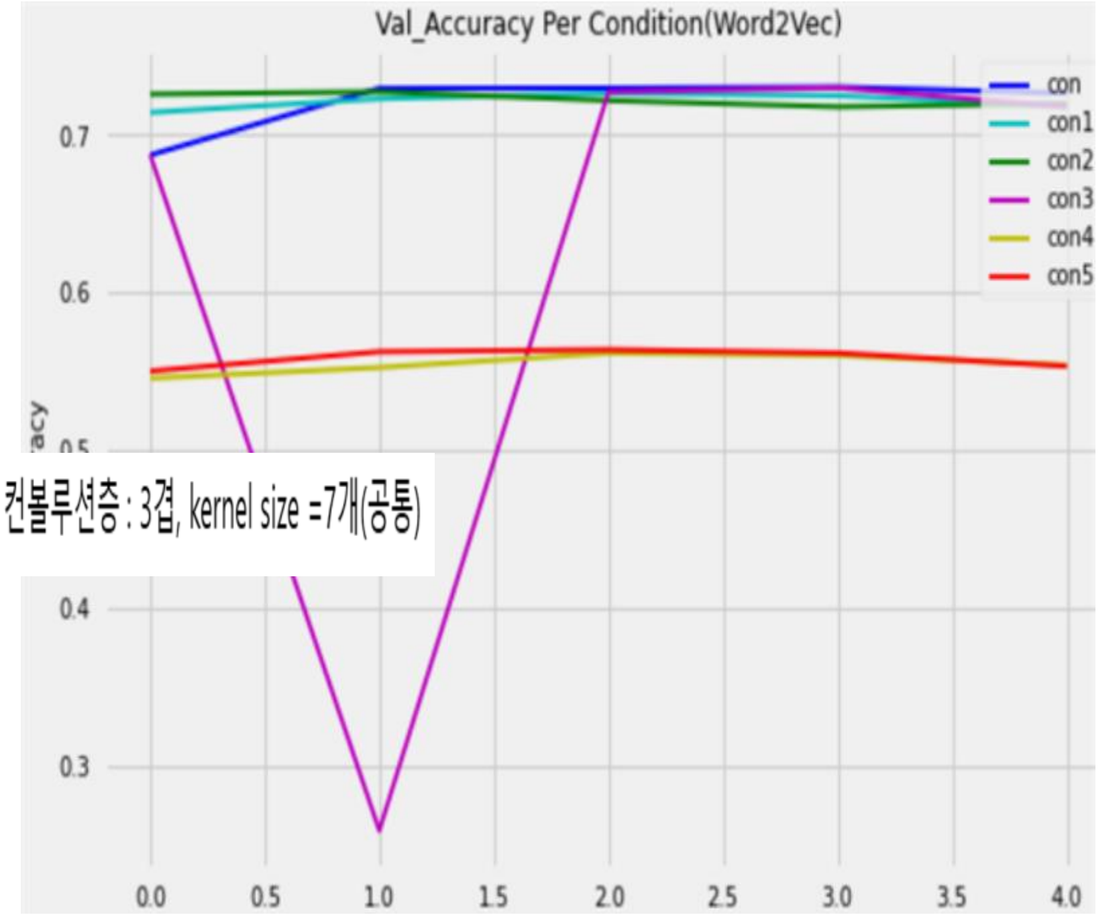
```
Epoch 1/20
1372/1372 - 138s - loss: 1.3390 - accuracy: 0.4189 - val_loss: 1.0211 - val_accuracy: 0.5910
Epoch 2/20
1372/1372 - 111s - loss: 0.8842 - accuracy: 0.6608 - val_loss: 0.8517 - val_accuracy: 0.6816
Epoch 3/20
1372/1372 - 108s - loss: 0.6497 - accuracy: 0.7630 - val_loss: 0.8116 - val_accuracy: 0.7060
Epoch 4/20
1372/1372 - 107s - loss: 0.5140 - accuracy: 0.8149 - val_loss: 0.8409 - val_accuracy: 0.7111
Epoch 5/20
1372/1372 - 106s - loss: 0.4300 - accuracy: 0.8461 - val_loss: 0.8851 - val_accuracy: 0.7148
Epoch 6/20
1372/1372 - 106s - loss: 0.3696 - accuracy: 0.8696 - val_loss: 0.9576 - val_accuracy: 0.7157
Restoring model weights from the end of the best epoch.
Epoch 00006: early stopping
```

모델링 - Word2Vec + CNN (개선)

1. 파라미터 크게 (max_length = 500)
2. 컨볼루션 층 얇게 (컨볼루션층 = 2겹)
3. 커널 크기를 작게 (kernel = 3개 - 공통)
4. 파라미터 작게 (vocab_size = 500)
5. 파라미터 작게 (vocab_size = 500) + 컨볼루션층 얇게 (컨볼루션층 = 2겹)

모델링 - Word2Vec + CNN (개선)



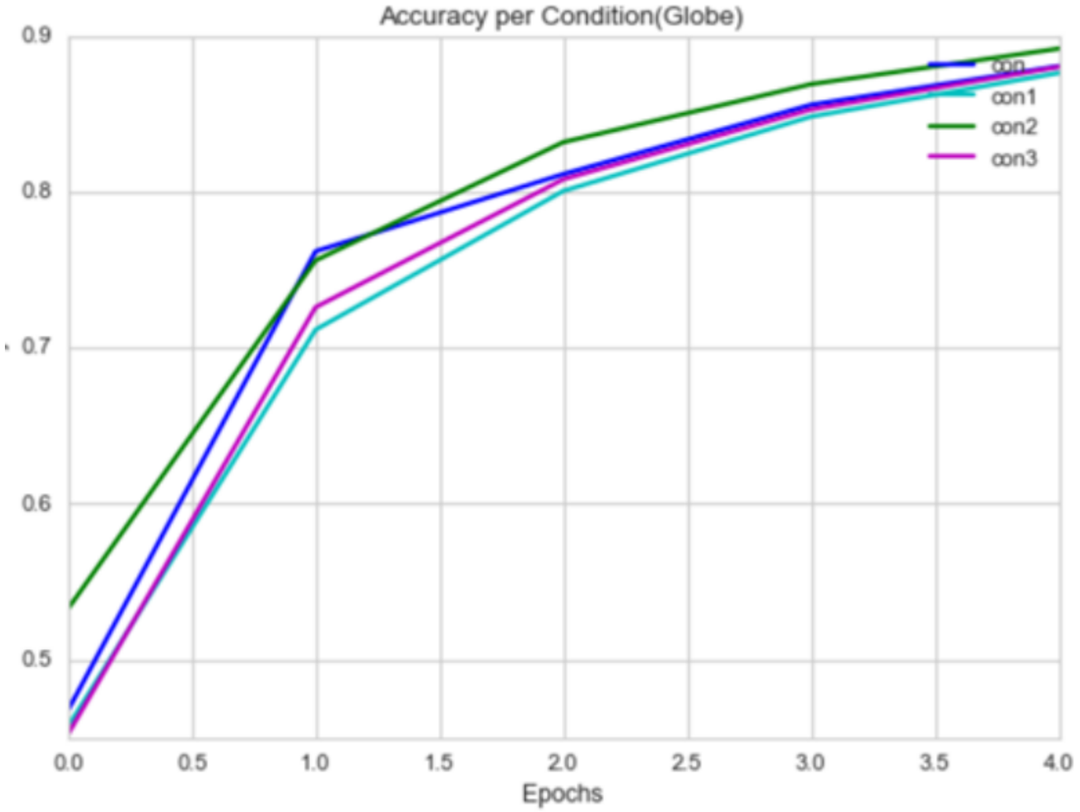


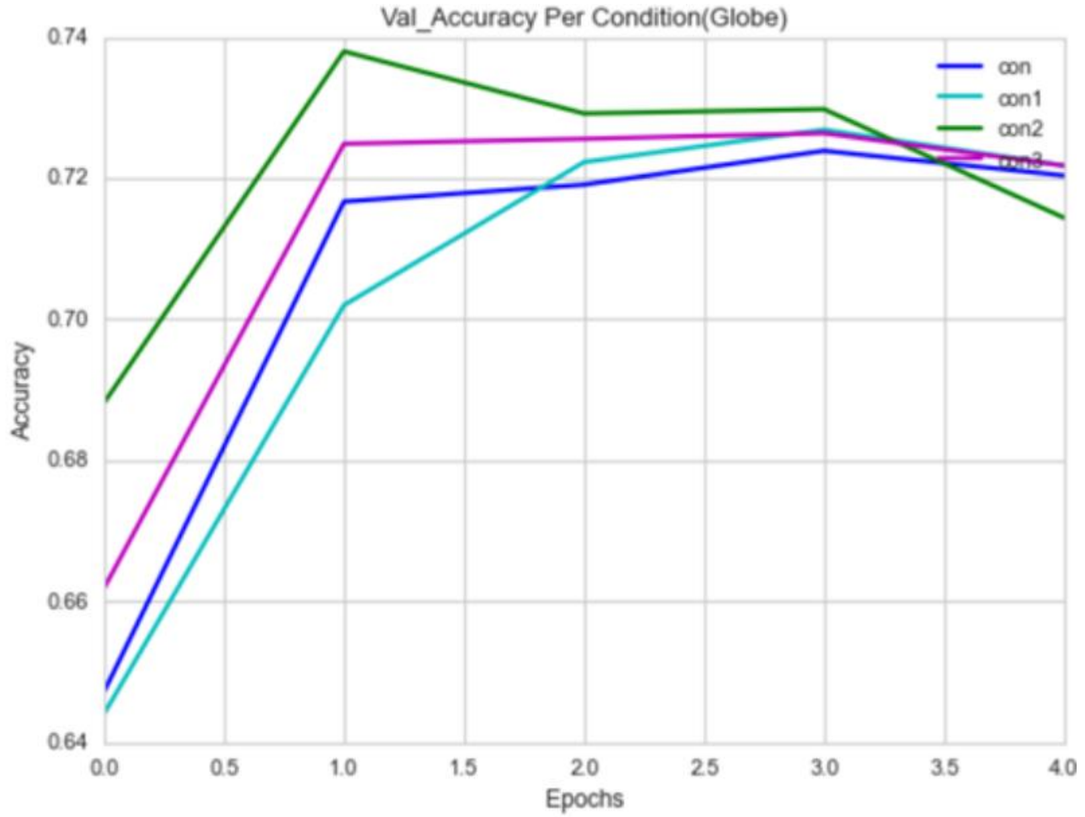
vocab_size = 53698, max_legnth = 212, 컨볼루션층 : 3겹, kernel size =7개(공통)

모델링 - Glove + CNN (개선)

1. 파라미터 크게 (max_length = 500)
2. 컨볼루션 층 얇게 (컨볼루션층 = 2겹)
3. 커널 크기를 작게 (kernel = 3개 공통)

모델링 - Glove + CNN (개선)





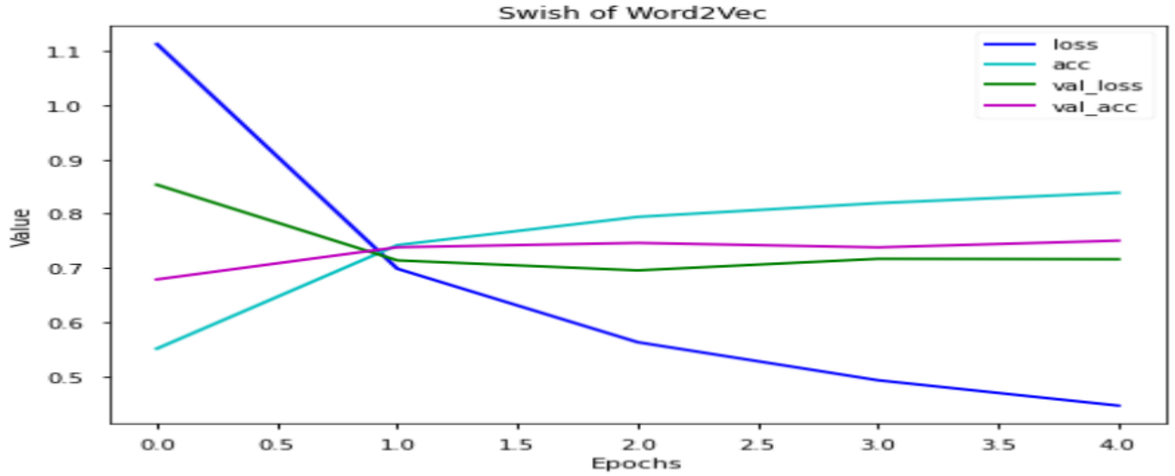
모델링 - Word2vec + BI-LSTM

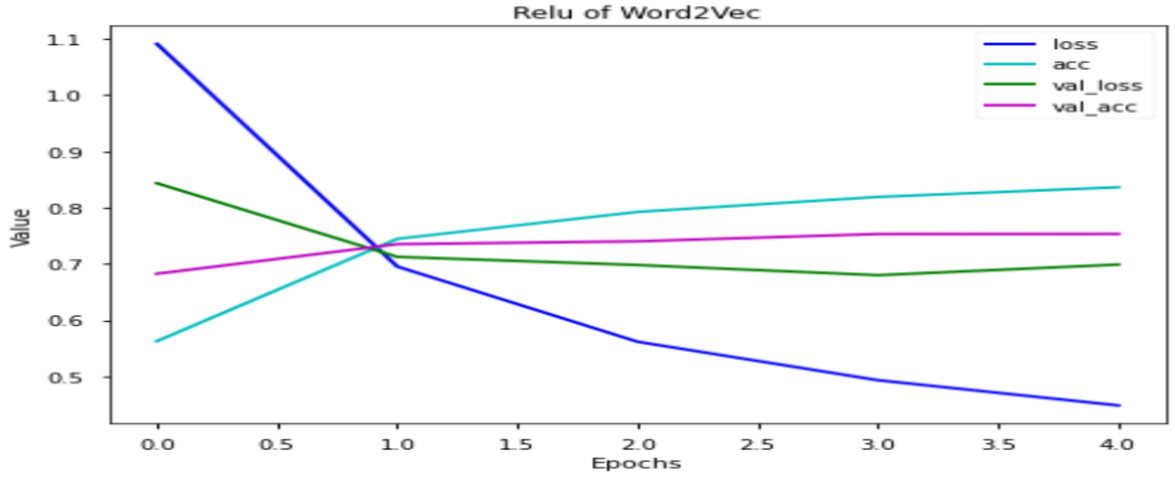
```
Epoch 1/20
1372/1372 - 312s - loss: 1.1128 - accuracy: 0.5511 - val_loss: 0.8533 - val_accuracy: 0.6788
Epoch 2/20
1372/1372 - 308s - loss: 0.6985 - accuracy: 0.7420 - val_loss: 0.7138 - val_accuracy: 0.7380
Epoch 3/20
1372/1372 - 306s - loss: 0.5629 - accuracy: 0.7940 - val_loss: 0.6954 - val_accuracy: 0.7460
Epoch 4/20
1372/1372 - 304s - loss: 0.4927 - accuracy: 0.8194 - val_loss: 0.7167 - val_accuracy: 0.7378
Epoch 5/20
1372/1372 - 306s - loss: 0.4460 - accuracy: 0.8386 - val_loss: 0.7160 - val_accuracy: 0.7502
Epoch 6/20
1372/1372 - 304s - loss: 0.4121 - accuracy: 0.8507 - val_loss: 0.7263 - val_accuracy: 0.7433
Restoring model weights from the end of the best epoch.
Epoch 00006: early stopping
```

모델링 - Glove + BI-LSTM

```
Epoch 1/20
1372/1372 - 194s - loss: 1.1428 - accuracy: 0.5432 - val_loss: 0.9091 - val_accuracy: 0.6532
Epoch 2/20
1372/1372 - 188s - loss: 0.7549 - accuracy: 0.7176 - val_loss: 0.7730 - val_accuracy: 0.7122
Epoch 3/20
1372/1372 - 185s - loss: 0.6249 - accuracy: 0.7709 - val_loss: 0.7243 - val_accuracy: 0.7311
Epoch 4/20
1372/1372 - 183s - loss: 0.5461 - accuracy: 0.8002 - val_loss: 0.7136 - val_accuracy: 0.7392
Epoch 5/20
1372/1372 - 184s - loss: 0.4929 - accuracy: 0.8203 - val_loss: 0.7128 - val_accuracy: 0.7392
Epoch 6/20
1372/1372 - 186s - loss: 0.4547 - accuracy: 0.8353 - val_loss: 0.7125 - val_accuracy: 0.7476
Epoch 7/20
1372/1372 - 188s - loss: 0.4251 - accuracy: 0.8462 - val_loss: 0.7218 - val_accuracy: 0.7469
Epoch 8/20
1372/1372 - 186s - loss: 0.4015 - accuracy: 0.8555 - val_loss: 0.7377 - val_accuracy: 0.7444
Epoch 9/20
1372/1372 - 186s - loss: 0.3827 - accuracy: 0.8629 - val_loss: 0.7539 - val_accuracy: 0.7416
Restoring model weights from the end of the best epoch.
Epoch 00009: early stopping
```

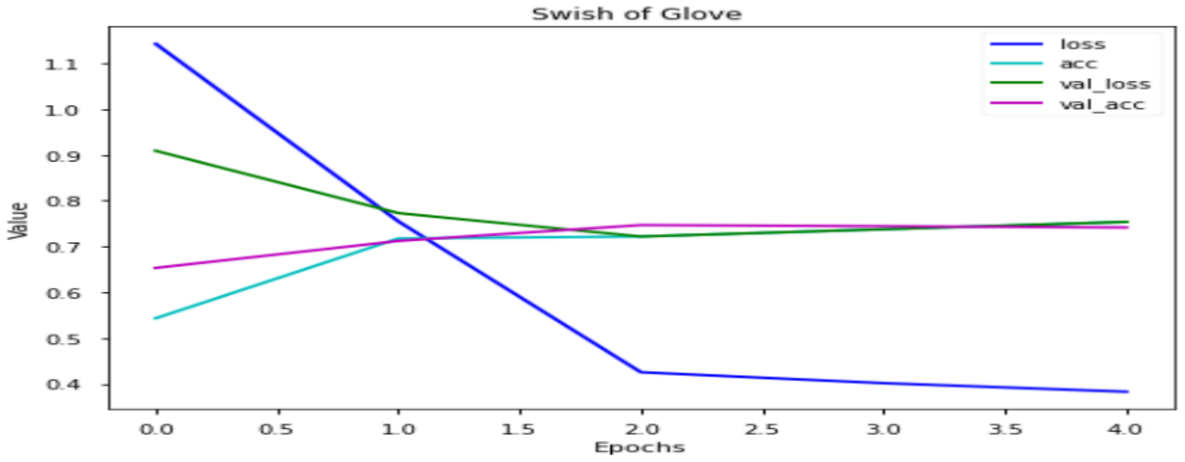
모델링 - Word2vec + BI-LSTM (개선)

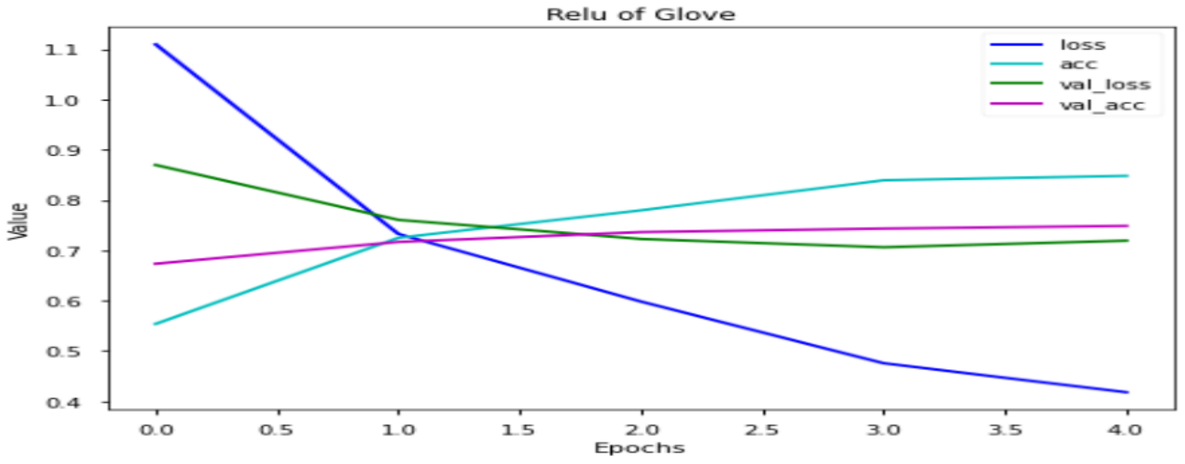




활성화 함수를 변경
(activation)
swish VS relu

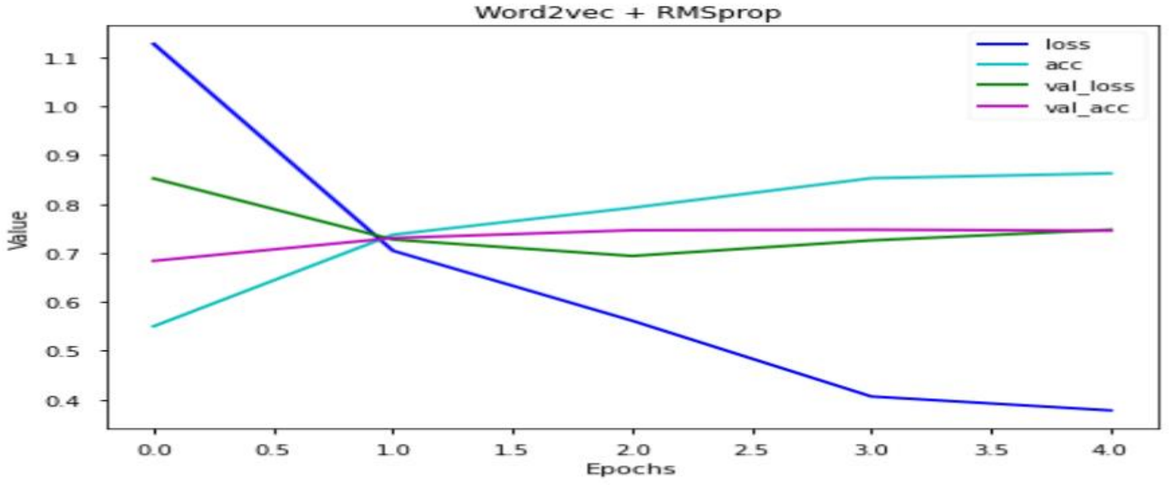
모델링 - Glove + BI-LSTM (개선)

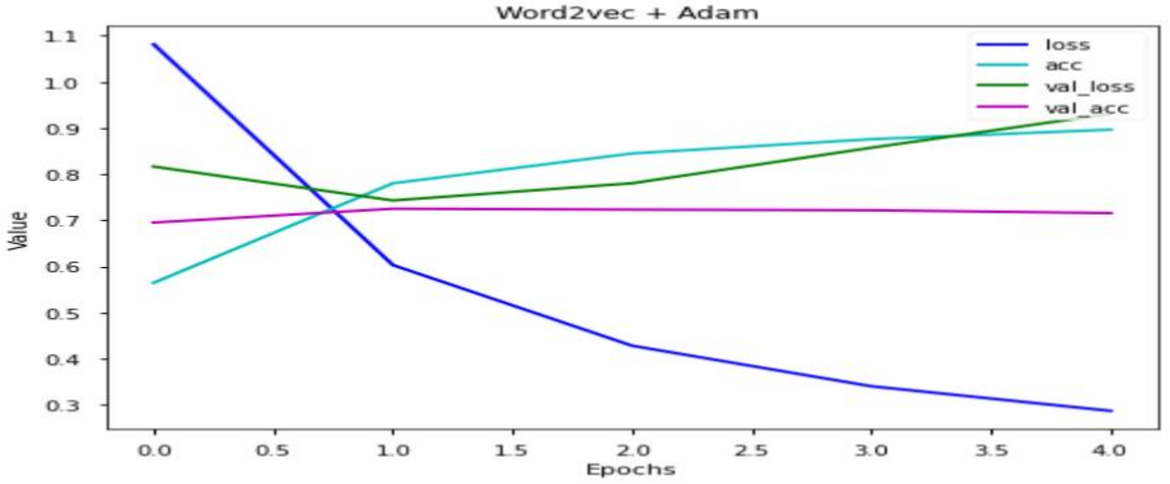




활성화 함수를 변경
(activation)
swish VS relu

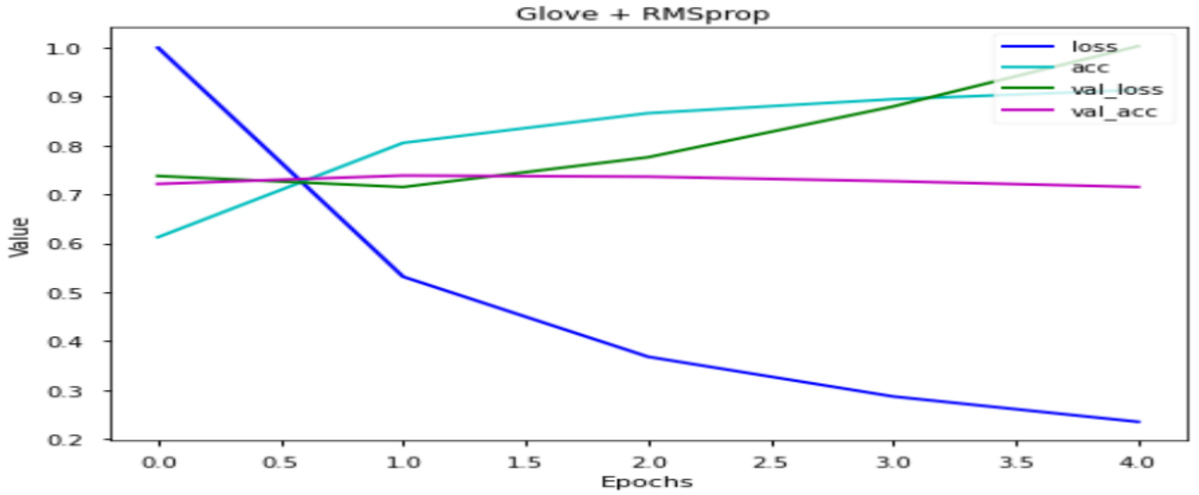
모델링 - Word2vec + BI-LSTM (개선)

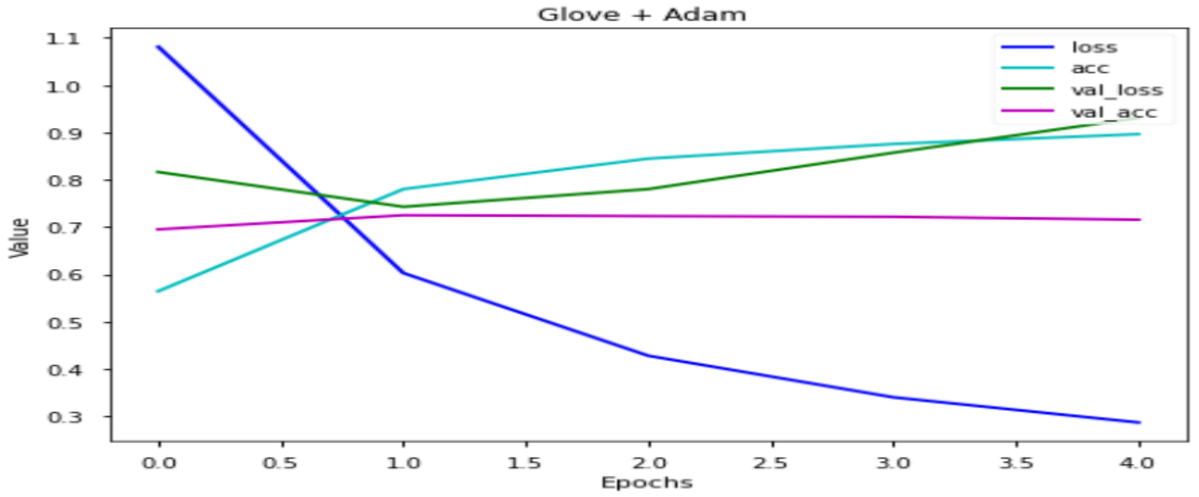




Optimizer를 변경
RMSprop VS Adam

모델링 - Glove + BI-LSTM (개선)





Optimizer를 변경
RMSprop VS Adam

모델링 - Word2vec + BI-LSTM (개선)

activation = swish
optimizer = RMSprop
learning rate = 0.001

모델링 - Word2vec + BI-LSTM (개선)

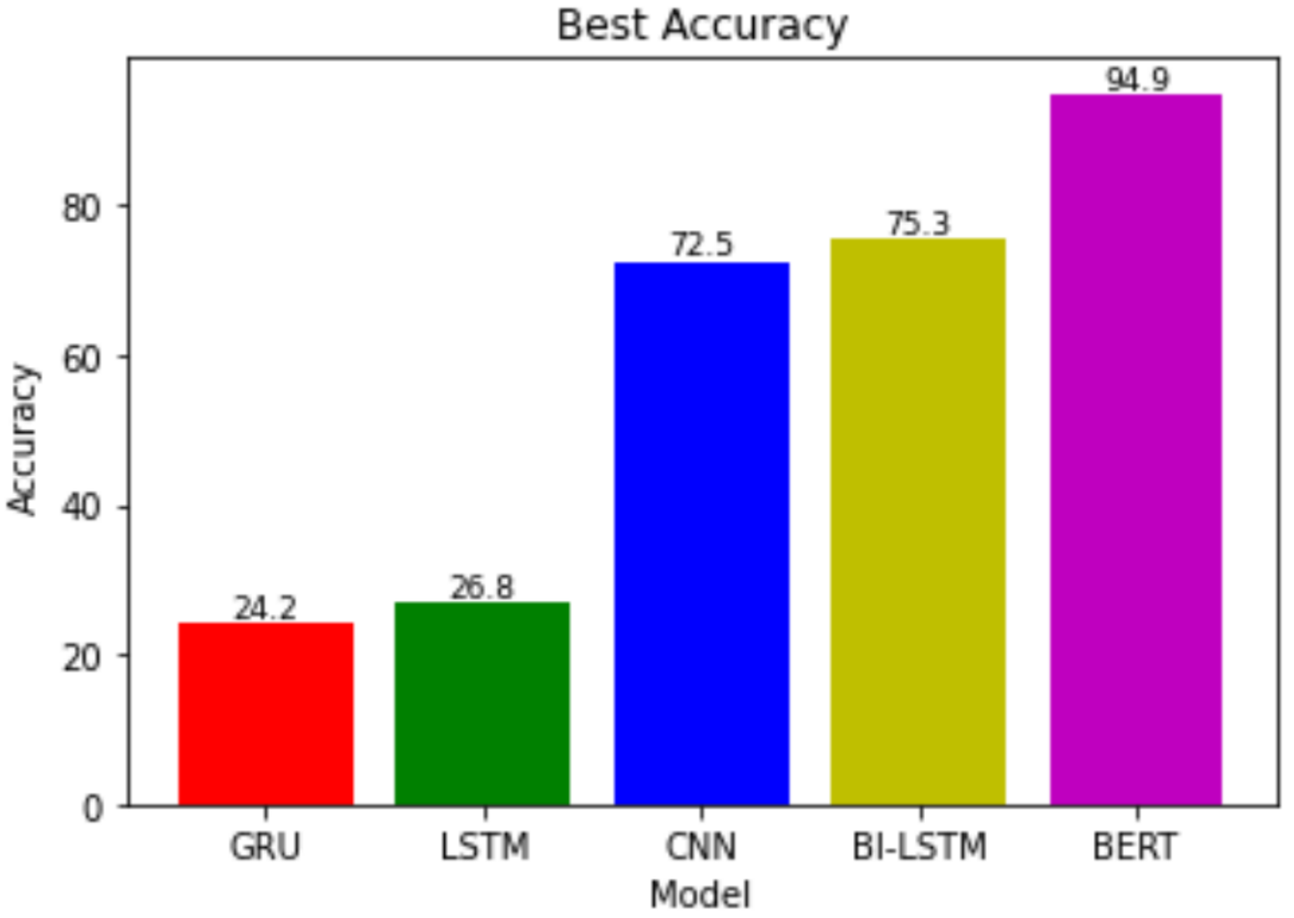
```
Epoch 1/20
1372/1372 - 311s - loss: 1.0910 - accuracy: 0.5628 - val_loss: 0.8434 - val_accuracy: 0.6825
Epoch 2/20
1372/1372 - 310s - loss: 0.6953 - accuracy: 0.7446 - val_loss: 0.7125 - val_accuracy: 0.7351
Epoch 3/20
1372/1372 - 307s - loss: 0.5617 - accuracy: 0.7925 - val_loss: 0.6980 - val_accuracy: 0.7402
Epoch 4/20
1372/1372 - 308s - loss: 0.4933 - accuracy: 0.8191 - val_loss: 0.6799 - val_accuracy: 0.7532
Epoch 5/20
1372/1372 - 306s - loss: 0.4484 - accuracy: 0.8363 - val_loss: 0.6988 - val_accuracy: 0.7533
Epoch 6/20
1372/1372 - 306s - loss: 0.4135 - accuracy: 0.8500 - val_loss: 0.7268 - val_accuracy: 0.7477
Epoch 7/20
1372/1372 - 307s - loss: 0.3859 - accuracy: 0.8604 - val_loss: 0.7610 - val_accuracy: 0.7420
Restoring model weights from the end of the best epoch.
Epoch 00007: early stopping
```

모델링 - Bert

```
Epoch 1/10
5488/5488 [==============================] - 3334s 602ms/step - loss: 0.7885 - accuracy: 0.6975 - val_loss: 0.4384 - val_accuracy: 0.8370
Epoch 2/10
5488/5488 [==============================] - 3302s 602ms/step - loss: 0.2576 - accuracy: 0.9082 - val_loss: 0.4592 - val_accuracy: 0.8386
Epoch 3/10
5488/5488 [==============================] - 3299s 601ms/step - loss: 0.0792 - accuracy: 0.9729 - val_loss: 0.4801 - val_accuracy: 0.8580
Restoring model weights from the end of the best epoch.
Epoch 00003: early stopping
training model for CV #2
Epoch 1/10
5488/5488 [==============================] - 3318s 602ms/step - loss: 0.3740 - accuracy: 0.8689 - val_loss: 0.3381 - val_accuracy: 0.8736
Epoch 2/10
5488/5488 [==============================] - 3301s 602ms/step - loss: 0.1139 - accuracy: 0.9615 - val_loss: 0.2761 - val_accuracy: 0.9083
Epoch 3/10
5488/5488 [==============================] - 3307s 603ms/step - loss: 0.0446 - accuracy: 0.9849 - val_loss: 0.3311 - val_accuracy: 0.9091
Restoring model weights from the end of the best epoch.
Epoch 00003: early stopping
training model for CV #3
Epoch 1/10
5488/5488 [==============================] - 3315s 602ms/step - loss: 0.2151 - accuracy: 0.9288 - val_loss: 0.1176 - val_accuracy: 0.9596
Epoch 2/10
5488/5488 [==============================] - 3301s 601ms/step - loss: 0.0617 - accuracy: 0.9802 - val_loss: 0.1331 - val_accuracy: 0.9573
Epoch 3/10
5488/5488 [==============================] - 3297s 601ms/step - loss: 0.0380 - accuracy: 0.9880 - val_loss: 0.1872 - val_accuracy: 0.9447
Epoch 4/10
5488/5488 [==============================] - 3297s 601ms/step - loss: 0.0290 - accuracy: 0.9906 - val_loss: 0.1847 - val_accuracy: 0.9498
Epoch 5/10
1593/5488 [=======>......................] - ETA: 36:04 - loss: 0.0193 - accuracy: 0.9930
```

모델링 - 모델 별 Best Accuracy

Model	Best Accuracy
GRU	0.2427
LSTM	0.2680
CNN	0.7259
BI-LSTM	0.7533
Bert	0.9573



피드백


Q&A

소감한마디

곽희원 공부했던 NLP를 직접 구현하면서 배우는 뜻깊은 시간이었습니다. 이번 팀 미니 프로젝트를 진행하면서 제가 하고 싶었던 여러 과정을 실행할 수 있게 해주셔서 감사합니다. 또한, 발표자로서 맘 편하게 발표하게 열심히 PPT 만들어 주시고 열심히 노력해주셔서 감사합니다.

김홍범 이번 대회를 진행하면서 팀별 단위의 프로젝트의 장점을 깨달았습니다. 1가지 모델이라도 정확성을 올리기 위한 방법들은 정말 다양하였고 이를 혼자 수행했다면 정말 긴 시간이 걸렸을 것입니다. 또한 EDA과정부터 많은 것들을 시도해보고 싶었지만 시간의 한계로 인해 많이 못해서 정말 아쉬웠고 혼자라면 못했을 결과를 함께 이루어낸 거 같아 팀원들에게 너무 고맙습니다~발표 전날까지 다들 고생하셨습니다:)

선우미 NLP를 이 수업을 통해 처음 접하게 되었는데,이론에서부터 실제 코드 작성까지 해보니 더욱 흥미가 생겼습니다.
무엇보다도 좋은 팀원분들을 만나 협력하는 경험을 가질 수 있게 해주셔서 감사합니다.

이원권 기존에 막연하게만 생각했던 NLP 프로젝트를 진행하면서 관심도 갖게 되었고 직접 전처리부터 모델링까지 구축하면서 많은 Hands-on 경험을 얻었습니다. 모델 구성이나 하이퍼 파라미터, 그리고 활성화 함수나 옵티마이저까지 다양한 세팅에서 실험을 하면서 조금씩 정확도가 오르는 것을 보면서 정말 인상 깊었고 다음에는 BERT 모델까지 구현해보고
싶다고 느꼈습니다. 마지막으로 주말부터 제출 전 날 늦은 밤까지 고생한 조원들 모두 고생했고 수고하셨습니다!

정혜원 이번 프로젝트는 배운 것을 직접 해보며 더 깊이 이해할 수 있었던 시간이었습니다. 높은 정확도를 내기 위해 여러 시도의 노력이 필요하다는 걸 알게 되었고 생각대로 구현할 수 없는 아쉬움은 있었지만 열정적인 팀원들을 보고 힘을 낼 수 있었습니다. 모두 수고 많으셨어요~

마지막으로

발표를 들어주셔서 감사합니다