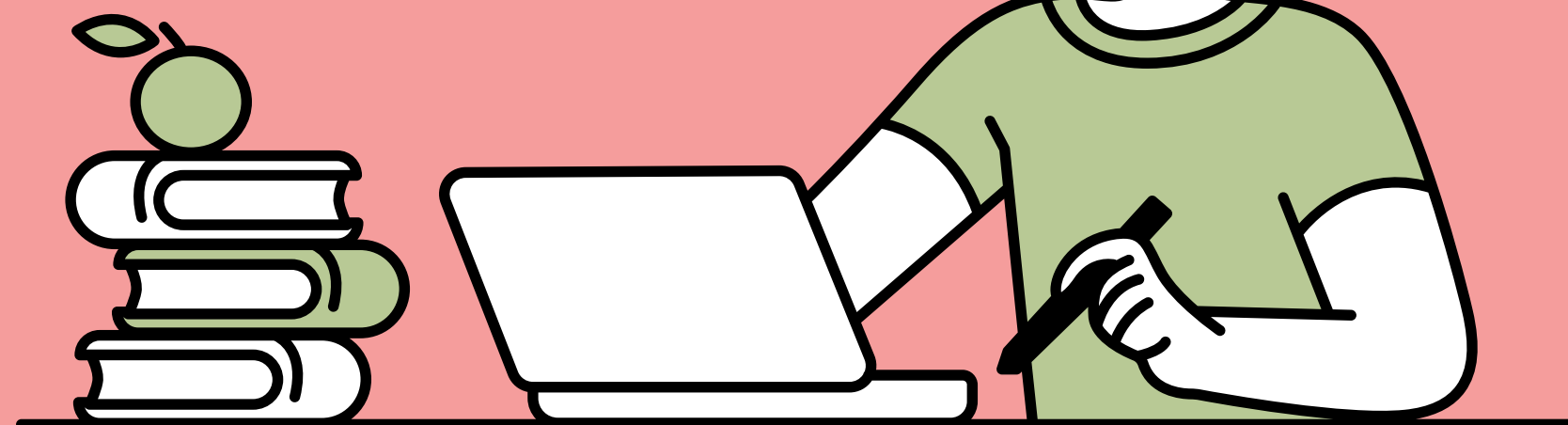


나만의 책 제목 만들기 & 책 추천 서비스

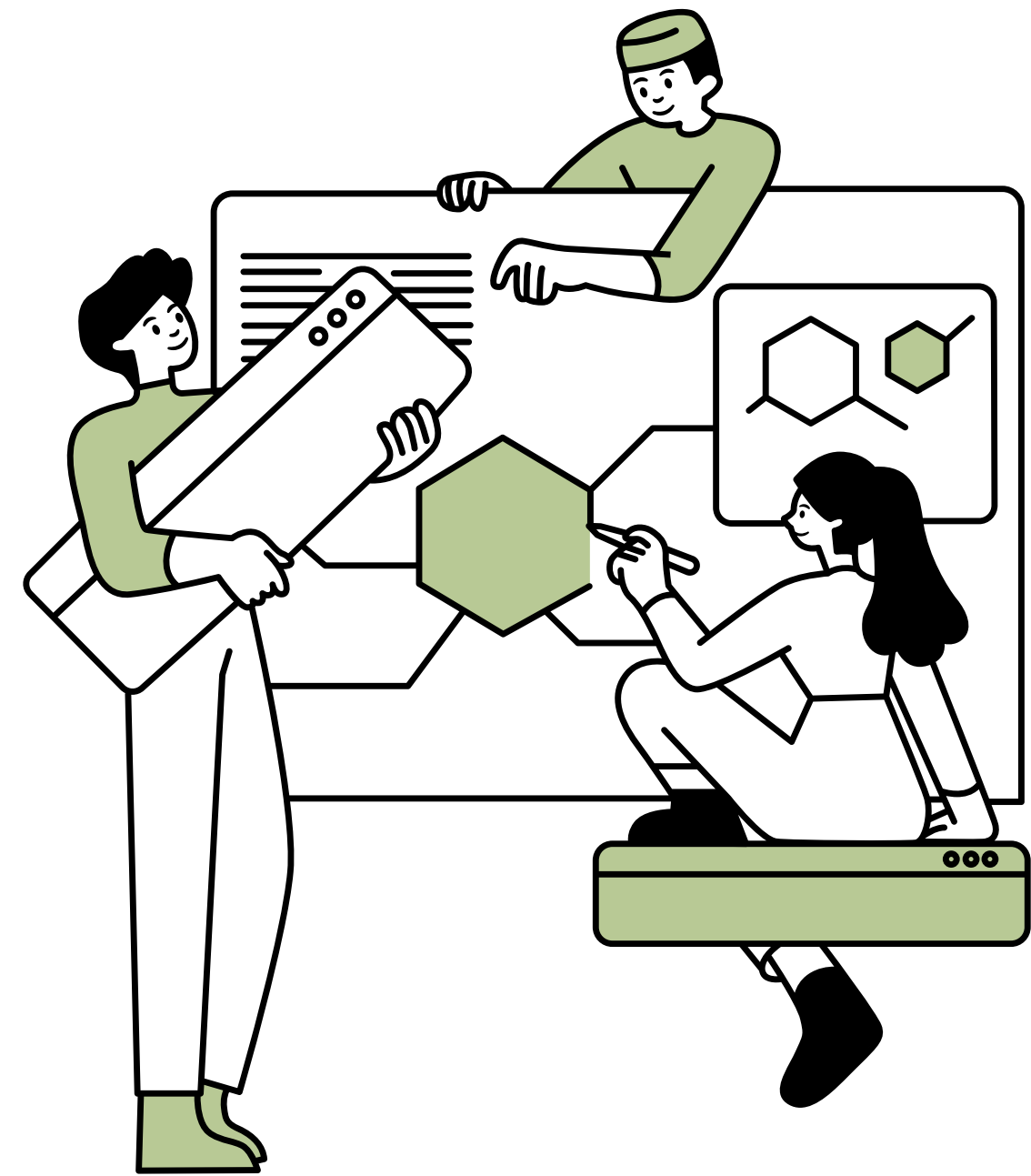


코드스테이츠
AI 15기 전현아



목차

- 01 주제 선정 이유
- 02 데이터 전처리
- 03 딥러닝 적용 & 모델 예측
- 04 책 추천 서비스
- 05 프로젝트 회고





책 제목 생성 모델을 만든 이유

저자들 대부분이 책 제목 만드는 것을 버거워 합니다.
일상생활에서도 백일장, 시 짓기 등 제목을 만들어야 하는 경우가 있습니다.
그래서 책 요약본을 입력하면 내용에 맞는 제목을 만들어주는 모델을 만들고 싶었습니다.



데이터

AI hub - 인기 대출 도서(4개 csv)

책 제목, 책 요약본 사용

	BOOK_TITLE_NM	AUTHR_NM	BOOK_INTRCN_CN
0	김상욱의 양자 공부 :완전히 새로운 현대 물리학 입문	지은이: 김상욱	양자 역학의 탄생부터 최전선까지 모든 역사와 20세기 물리학의 '제자백가'들이 펼치...
1	눈물을 마시는 새	이영도 지음	로 한국 판타지소설계의 정상에 선 이영도의 신작이 출간됐다. 이전 작품들에서 한 걸...
2	나의 라임 오렌지나무	J. M. 데 바스콘셀로스	우리 나라 독자들에게 많은 사랑을 받아온 가 완역본 보급판으로 재출간됐다. 100여...
3	(N잡하는 허대리의) 월급 독립 스쿨 :돈 버는 재미를 잃어버린 직장인을 위한	지은이: N잡하는 허대리	평범한 직장인에서 월급 외 수익 1,000만 원을 달성한 동시에 'N잡하는 허대리'...
4	숫! 신데렐라는 시계를 못 본대 :길이 재기와 시계 보기	글: 고자현 ;그림: 김명곤	초등 1.2학년 수학동화 시리즈 1권. '길이 재기와 시계 보기(측정)'에 관련된 ...
...
27489	Gorilla (고릴라)	Anthony Browne	너무나 바쁜 아빠 대신 아빠가 사준 고릴라 인형과 즐거운 한 때를 보내는 한나. 섬...
27490	그림책의 책 :100개의 주제로 엮은 그림책 북큐레이션 북	제님 지음	15년간 기록한 1만여 권의 목록에서 1300여 권의 그림책을 엄선하여 100개의 ...
27491	걸리버 여행기	조나단 스위프트 원작;윤유리 그림	Part 1에서는 우선 명작을 패러디한 만화로 작품의 줄거리, 주제를 한눈에 살펴본...
27492	가족을 지켜라!	김바다 (지은이), 국은오 (그림)	다른 집들과 마찬가지로 지우네도 코로나19로 인하여 외출도 외식도 한 지 오래였다....
27493	이게 뭐예요? :자연에서 볼 수 있는 온갖 모양들	글: 라파엘 마르탱 ;옮김: 강현주	자연에 관심이 많은 작가 라파엘 마르탱은 이 책을 통해 평소 우리 주변에서 흔히 볼...

27494 rows × 3 columns

데이터 전처리

vocab에 네이버 리뷰 단어를 저장 후
책제목, 요약 토큰화

0	
0	[PAD]
1	[UNK]
2	[CLS]
3	[SEP]
4	[MASK]
...	
29995	말밖에는
29996	맘을
29997	맛도
29998	망하지
29999	망한다

30000 rows × 1 columns

seq2seq를 위해 decoder input, target을 나눠줌

decoder_input	decoder_target
[PAD] 김상욱의 양자 공부 :완전히 새로운 현대 물리학 입문	김상욱의 양자 공부 :완전히 새로운 현대 물리학 입문 [UNK]
[PAD] 눈물을 마시는 새	눈물을 마시는 새 [UNK]
[PAD] 나의 라임 오렌지나무	나의 라임 오렌지나무 [UNK]
[PAD] (N잡하는 허대리의) 월급 독립 스쿨 :돈 버는 재미를 잃어버린 직장인을...	(N잡하는 허대리의) 월급 독립 스쿨 :돈 버는 재미를 잃어버린 직장인을 위한 [...]
[PAD] 씯! 신데렐라는 시계를 못 본대 :길이 재기와 시계 보기	씯! 신데렐라는 시계를 못 본대 :길이 재기와 시계 보기 [UNK]

토큰화 후 임베딩

```
array([[17857, 2269, 24600, ..., 0, 0, 0],
       [ 8335, 18869, 1829, ..., 0, 0, 0],
       [ 6829, 1290, 2348, ..., 0, 0, 0],
       ...,
       [23634, 1610, 8658, ..., 0, 0, 0],
       [11747, 8550, 1290, ..., 0, 0, 0],
       [ 5933, 1537, 2211, ..., 0, 0, 0]])
```

딥러닝

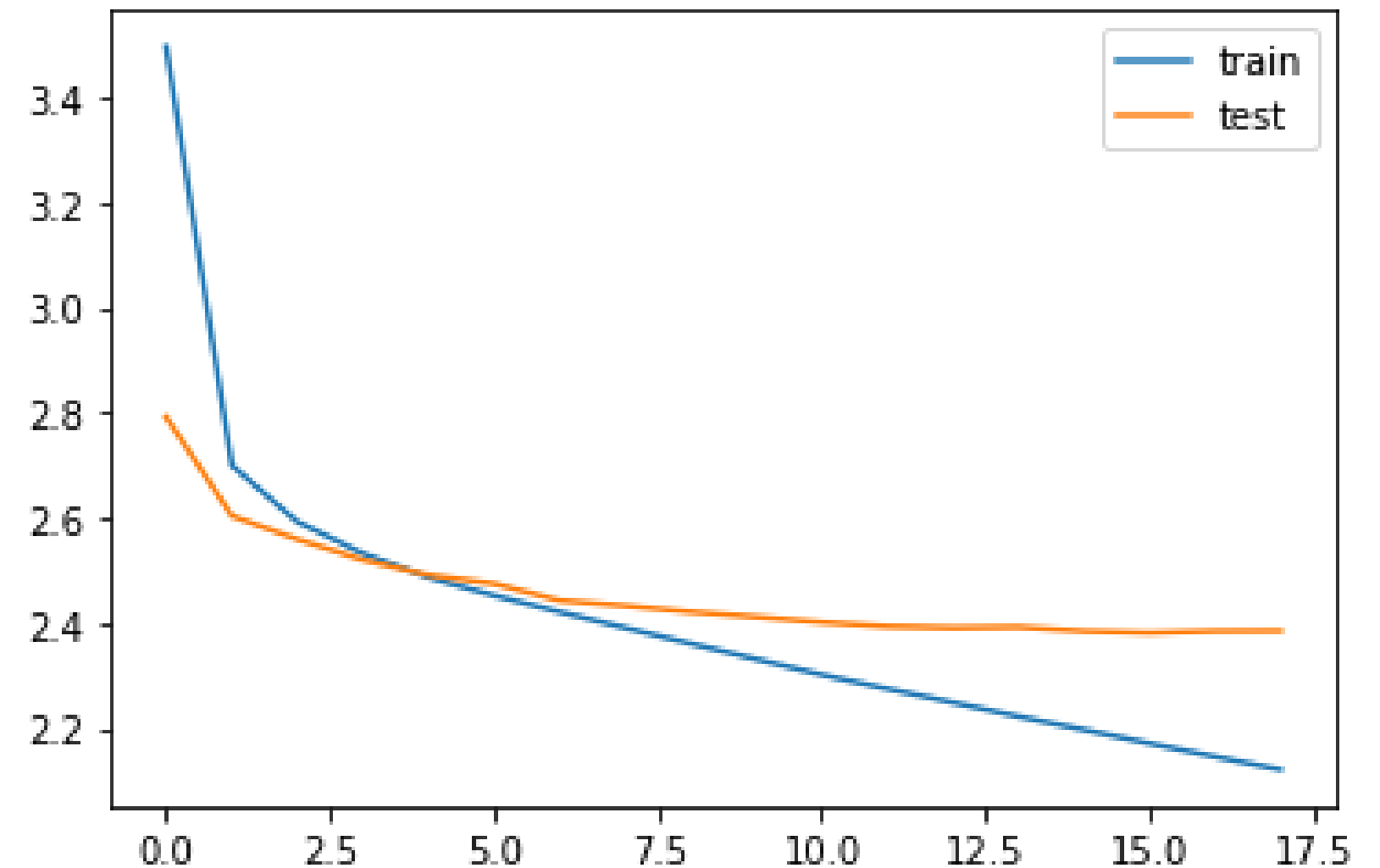
seq2seq+Attention모델

인코더: LSTM 3개/4개

디코더: LSTM 1개

어텐션: 미리 만들어진 어텐션 함수 사용

loss함수



모델 예측

(LSTM 3개)

원문 : 수학 주제 들을 수학 자들의 목소리를 통해 이야기 형식 으로 들려 주는 수학 자가 들려 주는 수학 이야기 시리즈] 수학 교과서 내용을 생동감 있게 재구성 었 으며 중간중간에 수학 자들의 에피소드 곁 들어 읽는 재미를 더(중략)

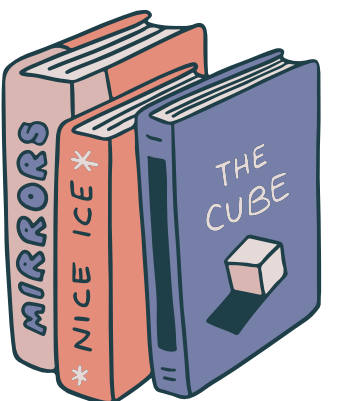
실제 제목 : PAD 푸 리에 들려 주는 삼각 함 이야기]]

예측 제목 : 수학

원문 : 초등학교 과학 교과서 실 린 개념 총 망 라 과학 사전 초등학생 과학 사 전 초등학교 교과서 지 도 서 나오는 개념 아이들이 이해하기 쉽 도록 엮 책 입니다] 지구 물질 생명 에너 지의 네 부분 으로 나누 어 가나다 순 으로 정리(중략)

실제 제목 : PAD 개념 쏙] 실력 쏙] 초등 학생 과학 사전]]

예측 제목 : 쿠 과학 과학 상식 학



모델 예측

원문 : y] 인 문 사회 학 습 만화 시리즈 권 1 자의 기원 탄생 자의 생 성 원 리와 부 수의 활용 등을 담
았다] 수많은 자의 기초 되는 기본 자와 자 쓰는 법 비슷 하거나 헛갈 리기 쉬운 자 다양한 자와 활용
법을 알려 준다 (중략)

실제 제목 : PAD W h y] UNK UNK 이야기]

예측 제목 : W h y y e i i i i i i i i i i i a a a a a b k e a i ## e U U

(LSTM4개)

원문 : 한국 주식 시장 지루한 박스 권을 형성 했던 이유 대세 상 승 언제 어떻게 오는 지 미국 달 러 한
국 주식 시장 어떻게 연결 되었 는지 도 업 종 어떻게 찾을 것인지 (중략)

실제 제목 : PAD 정 해진 미래 대세 상 승 장이 온다 이야기 꾀 시나리오 투자 법

예측 제목 : U 투자



모델 예측

원문 : 독일 게 프랑스 인 철학 연구 자 발 타 자 르 토마스 니 체의 허무 주의를 재조명 하며 고통 스러운 우리의 일상을 진 단한] 사색 하고 숙 고 하는 학 문 넘어서 철학을 우리의 삶에 적 용 해(중략)

실제 제목 : PAD 우울 할 땐 니 체

예측 제목 : 나는 산 U

원문 : FSILLYANDSLIGHTLYSCARYTRADITIONALTALLESANDSCHOOLYARDRHYMESFORBEGINNING R(중략)

실제 제목 : G ho st and O th er S ill y C h ill er s P ap er b ack

예측 제목 : D e e a a ##r a ##r ##ck ##ck e a ##e a ##r a t a ##r e e



책 추천 서비스

코사인 유사도를 이용해 비슷한 책 추천해주는 서비스



```
get_recommendation('나의 라임')
```

2

16382

21364

14362

18497

20022

12817

17554

4611

5410

16732

Name: BOOK_TITLE_NM, dtype: object

나의 라임 오렌지나무

밤 기차 여행

사랑의 역사 : 니콜 크라우스 장편소설

사냥꾼을 만난 꼬마곰

지금은 나만의 시간입니다 - 인생을 리셋하는 혼자 있는 시간의 힘

아빠 해마 이야기

난

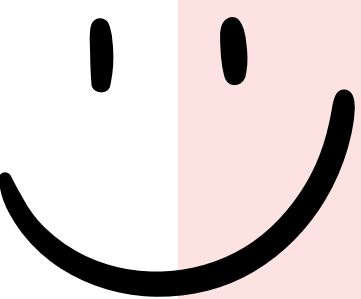
하루 5분, 명화를 읽는 시간 : 내 방에서 즐기는 반전 가득한 명화 이야기

(어린이를 위한) 우동 한 그릇

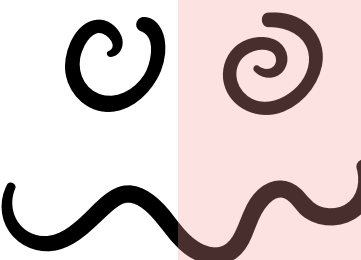
(그림책) 돌 씹어 먹는 아이

넌 어떻게 춤을 추니?

프로젝트 회고



- 섹션3 프로젝트와 연결지었고 저번에 하지 못한 책 추천 서비스를 구현했다.
- 자연어 처리를 위해 열심히 공부하고 노력했다.
- 프로젝트 하는 동안 열심히 집중해서 했고 결과가 나와서 흥미로웠다.



- 모델을 직접 만들었으면 좋았겠지만 아직 부족해 다른 사람이 쓴 모델을 사용해 아쉽고 능력이 안되어 제대로 예측하지 못해 아쉽다.
- vocab을 사용할 때 내가 가지고 있는 책 데이터로 만들었다면 예측을 더 잘 했을 거 같다. 시간과 공부가 충분했다면 가능할 거 같다.

감사합니다

코드스테이츠
AI 15기 전현아

