

Everytime Analyzer

에브리타임 강의평 기반 강의 추천



TEAM : YBIGTA

✓ 박준형

✓ 박지은

✓ 최서영

✓ 홍규원



ybigeee



목차

문제 정의

- 문제 정의
- 프로젝트 목표

로드맵

- 아키텍처
- 데이터셋
- 모델링
- FE/BE

결과물

- 시연
- 한계점

기타

- Q & A



ybigeee



문제 정의

피상적인 카테고리



실질적 니즈를 반영한
새로운 카테고리 제시

강의평을 일일이 확인하는 수고로움



사용자 쿼리 입력에 따른 강의 추천

카테고리의 병렬적 나열

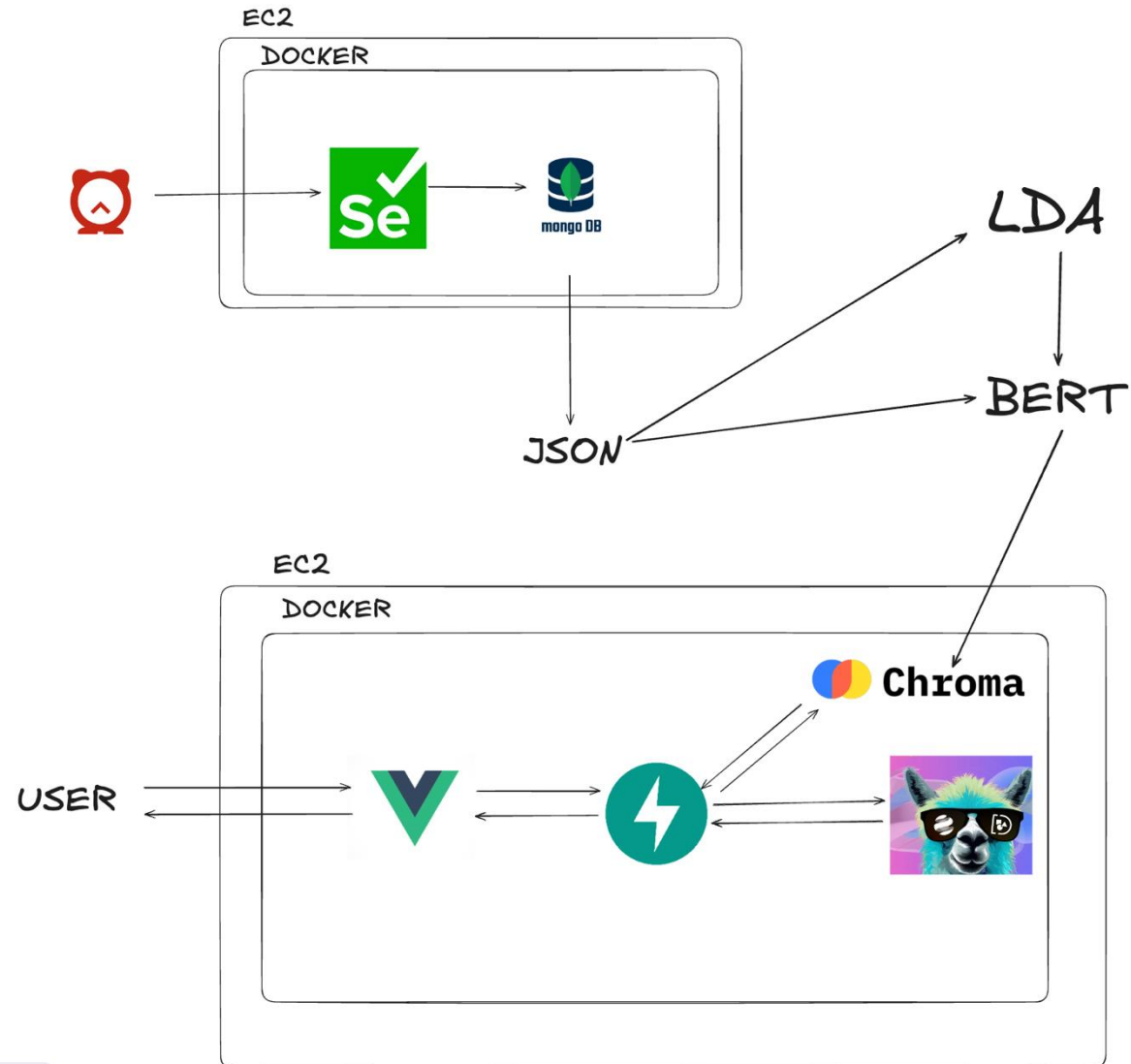


수강평에서 각 카테고리가
언급된 %를 제시

ybigeee



아키텍처



ybigeee



데이터 수집: 크롤링

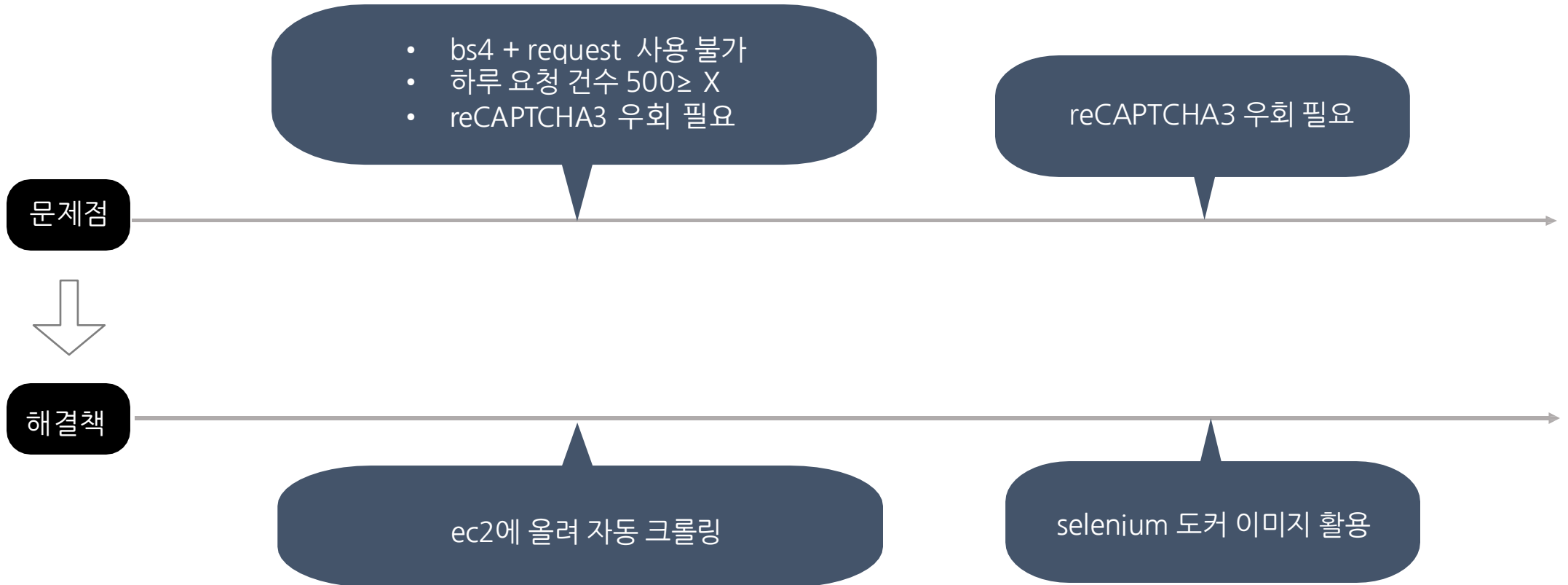
에브리타임 강의평 크롤링 w/ BS4, Kotlin

```
{
  _id: ObjectId('66b8f7aa593a1e64759b7ab4'),
  lectureCode: 'UCE1105-01-00',
  majorNth: 2,
  detailedNth: 1,
  lectureReview: {
    year: 2024,
    semester: '1학기',
    content: '최고의 교양..... 교수님 너무 친절하시고 좋음 수업도 재밌어요 교수님이 성적 잘주고싶어하심 출석은 녹강 1번(1시간) 놓쳤고 시험은 중간 2개 기
말 6개 틀림. A+받았어요. 족보 구해서 3개년치+필기본 약간만 외우시면 됩니다. 양이 많아서 미리 조금씩 해놓으면 좋을거같은데 전 그냥 벼락치기 했어요. 수업시
간에 중요하다고 하시는거는 꼭 나오는 것 같은데 저도 수업을 열심히 듣지는 않았습니다. 개인적으로 영화 좋아하는데 아는 영화 많이 나와서 중간중간 재미있었음.
출석방식은 그 전자출결 숫자 찍는걸로 합니다.',
    rate: 100
  }
},
{
  _id: ObjectId('66b8f7aa593a1e64759b7ab5'),
  lectureCode: 'UCE1105-01-00',
  majorNth: 2,
  detailedNth: 1,
  lectureReview: {
    year: 2024,
    semester: '1학기',
    content: '수업은 첫 수업 빼고 all 출퇴함 시험 전날 족보 3회독하면 a뽕 확정 ㅋㅋ 출석점수가 ㄹㅇ 크므로 결석은 하지 말자 참고로 난 중간 4개, 기말 2개
틀리고 대면수업 결석 1번 하고 a+받음',
    rate: 100
  }
}
```

ybigeee



데이터 수집: 크롤링에서의 troubleshooting





ybigeee



데이터셋

4.52 ★★★★★

전체 ▼ 등록순 ▼

★★★★★

추천 신고

24년 1학기 수강자

교수님 정말 착하시고 명강이에요! 강의력 정말 좋아요
하지만 꿀강인지는..ㅎㅎ

과제도 좀 많고 기말 시험 범위도 좀 방대하고, 개념이랑 관련된 실험
내용 위주로 좀 자세하게 다 외워야해서 쉽지는 않았어요

그래도 심리학 조금 관심 있으시다면 꼭 들어보세요!
이 수업 듣고 인사이드아웃 보면 진짜 더 많은 게 보입니다!!

★★★★★

추천 신고

24년 1학기 수강자 2

행겨야할 거 진짜 많습니다ㅠㅠ 하나 빠지면 성적x같이 나왔네
요.. 어려운 3천/4천 수학적 전공과목보다 성적이 못해요..)

★★★★★

추천 신고

24년 1학기 수강자 6

교수님의 강의 능력은 괜찮았지만, 명강이지만 꿀강의는 아니라고
생각하지 않아요. 전공 수업이 아니라 교양 수업이었는데, 많은 시간
과 노력을 들였음에도 불구하고 최종 성적이 별로였어요...특히 시험
때 암기해야 할 내용이 너무 많아서 세세한 부분까지 외워야 했어요..

★★★★★

추천 신고

24년 1학기 수강자 5

과제도 많고 그레йд도 급간도 좁아서 많이 힘들어요.. 동강의 메리
트가 없어진 그런..... 전공도 많이 들어서 기말 때 죽을뻔했네요....
그리고 주변에 재수강 한다는 사람 많습니다 ㅎㅎ.. 그냥 대면 꿀교양
들으시길.....

순번	희망	캠퍼스	학년	단위	학정번호-분반-실습				학점	신청	교과목 EN
					동일	개요	계획	마일			
3	♡ 희망	신촌	0	1000	YCG1111-01-00	📄	✍	📅	3	📝 신청	철학과방법
4	♡ 희망	신촌	0	1000	YCG1004-01-00	📄	✍	📅	3	📝 신청	인문학에길을묻다
5	♡ 희망	신촌	0	1000	YCG1201-01-00	📄	✍	📅	3	📝 신청	인간의역사와미래의인간

교과목 개요

본 과목은 다양한 학술적 주제를 인문학적 사고로 현실 문제를 연계하여 사유할 수 있는 실사구시적 안목을 함양하는 데에 둔다. 이와 더불어 다양한 학문 안에 기초로서 자리하는 인문학적 사고를 탐색해 보고자 한다. 본 과목의 수강을 통해 사회실천의 문제를 더불어 고민해 볼 수 있는 계기를 부여하며, '사회'현장에 대한 식견을 높일 수 있는 기회를 제공하게 될 것이다.

에브리타임 강의평

연세포털 수강 편람의 강의 개요

- 24년 2학기 신촌캠퍼스 교양, 문과대, 상경대, 경영대 이과대, 공과대, 생명시스템대, 인공지능융합대, 신과대, 사회과학대, 교육과학대, 생활과학대 743개 강의의 강의 개요 및 18406개 강의평

ybigeee



데이터셋

```
{
  "lectureCode": "AAI4101-01-00",
  "content": "강 대충대충임,.. 과제도 본인이 채점하기 귀찮으셨는지 캐글 대회 시키고 상위 10프로인가 안에 들면 만점임 (따로 시험 문제를 내지 않아도 되고 캐글이 자체적으로 점수(등수)를 매겨주니 교수자 입장에선 1석 2조) 사실 배우려고 들을 건데 교양 카테고리나 채우는 게
  "content_split": [
    "강 대충대충임,..",
    "과제도 본인이 채점하기 귀찮으셨는지 캐글 대회 시키고 상위 10프로인가 안에 들면 만점임",
    "(따로 시험 문제를 내지 않아도 되고 캐글이 자체적으로 점수(등수)를 매겨주니 교수자 입장에선 1석 2조) 사실 배우려고 들을 건데 교양 카테고리나 채우는 게
    "단 수업에서 다루는 내용이 많아서 선행지식 없으면 수업 따라가기 힘들 수도 있음."
  ],
  "content_normalized": [
    "그냥 대충대충임,..",
    "과제도 본인이 채점하기 귀찮으셨는지 캐글 대회 시키고 상위 10프로인가 안에 들면 만점임",
    "따로 시험 문제를 내지 않아도 되고 캐글이 자체적으로 점수 등수를 매겨주니 교수자 입장에선 일석이조 사실 배우려고 들을 건데 교양 카테고리나 채우는 게
    "단 수업에서 다루는 내용이 많아서 선행지식 없으면 수업 따라가기 힘들 수도 있음."
  ]
}
```

문장 단위
split

맞춤법
교정



ybigeee



모델링: LDA를 이용한 토픽 모델링 (토픽 추출)

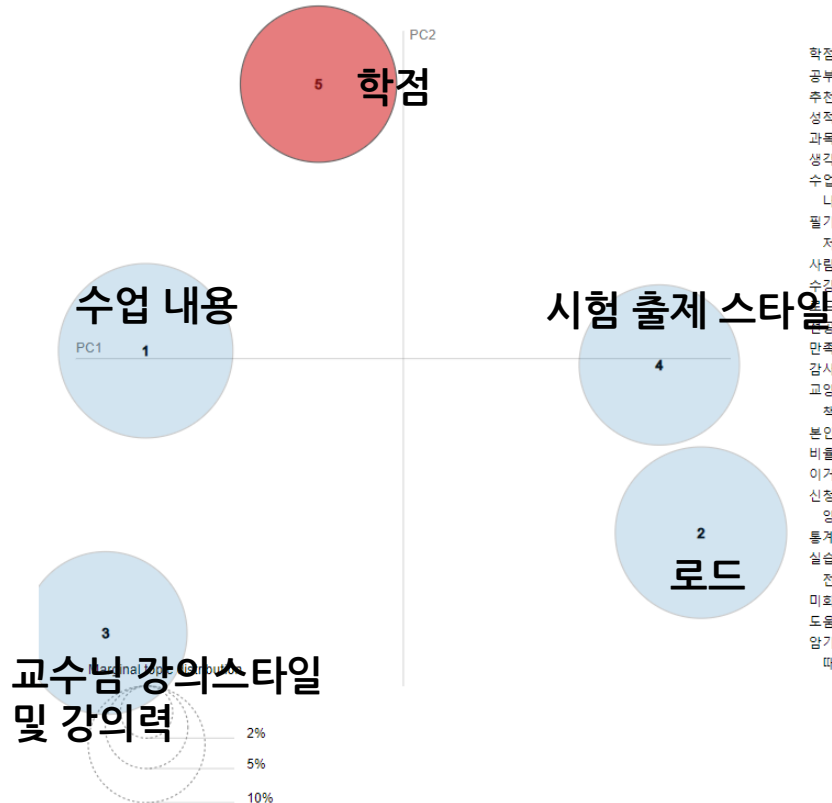
Selected Topic: Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾

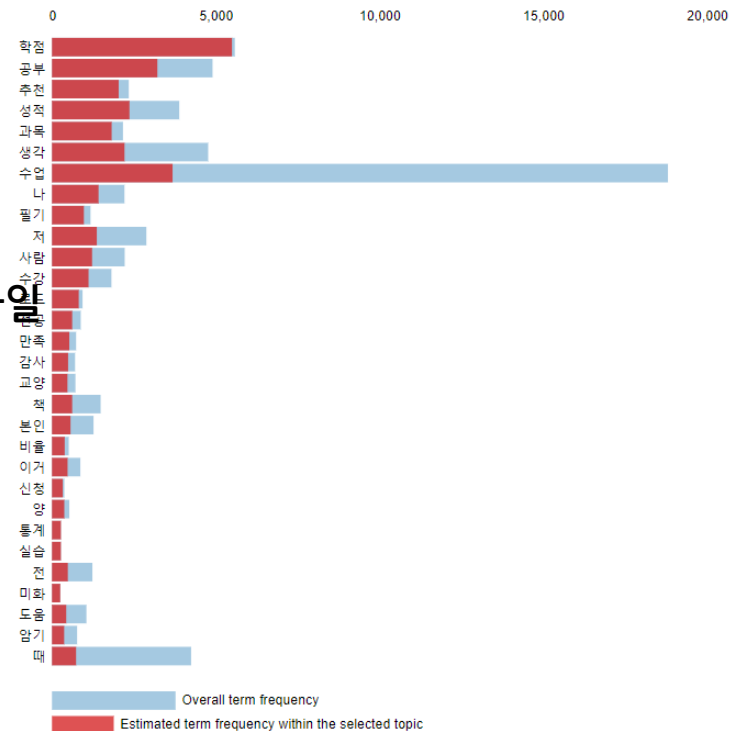
$\lambda = 0.6$

0.0 0.2 0.4 0.6 0.8 1.0

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 5 (17.9% of tokens)



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

* 토픽 별 상위 3개 키워드

1. 수업 내용: 수업, 내용, 강의
2. 로드: 과제, 기말, 중간
3. 교수님 강의스타일 및 강의력: 교수, 강의, 질문
4. 시험 출제 스타일: 문제, 시험, 개념
5. 학점: 학점, 공부, 추천

ybigeee



모델링: KoBERT를 이용한 토픽 라벨링 (분류)



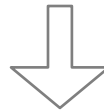
23년 1학기 수강자 2



교수님 강의력

지금까지 들었던 응통 전공 중에서 가장 좋았던 수업입니다. 교수님 설명도 굉장히 잘 해주시고 질문하면 잘 받아주십니다. 그리고 시험문제도 이번 학기는 4번 모두 너무 어렵지도, 쉽지도 않게 내셨고(100점 만점에 평균 60점 정도에서 왔다갔다 했습니다.) 학점도 후하게 주시는 것 같습니다. 홈워크 문제 열심히 풀어가시면 도움 될 것 같습니다.

학점



“교수님 설명도 굉장히 잘 해주시고 질문하면 잘 받아주십니다.”

교수님 강의력

“학점도 후하게 주시는 것 같습니다.”

학점



ybigeee



모델링: KoBERT를 이용한 토픽 라벨링(분류)



GPT 3.5 TURBO



Please check your email address for a confirmation link

Resend confirmation email



monologg/kobert



like 12



Feature Extraction



Transformers



PyTorch



JAX



Safetensors



Korean



bert



text-embeddings-inference



License: apache-2.0



Train



Deploy



Use this model

Step	Training Loss	Validation Loss	F1
100	1.596200	1.313232	0.483742
200	1.173400	1.072801	0.574504
300	0.938300	1.046649	0.588161
400	0.736900	1.057117	0.599950
500	0.583200	1.086699	0.601079
600	0.452900	1.170806	0.602579

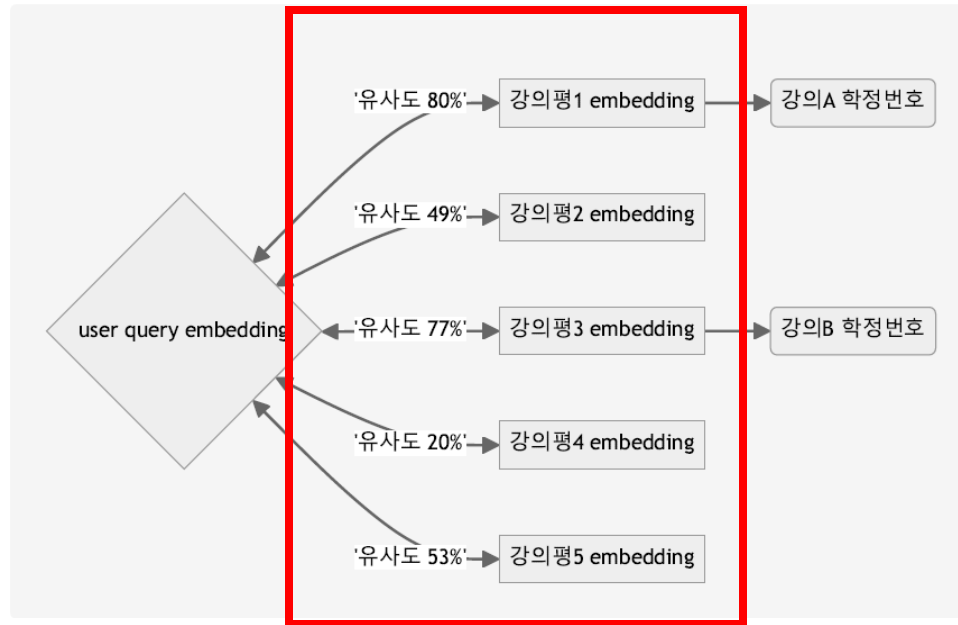
* KoBERT finetuning: 3 epoch로 제한



ybigeee



모델링: chroma 생성 및 데이터 적재



* 전체 강의평 문장의 embedding vector 생성 및 사용자 입력과의 유사도 계산

```

{
  "lectureCode": "AA12250-01-00",
  "content": "수업 운영이 주력구구식입니다. 중요 과제 공지도 사흘 전에 기말 과제가 계획에 있다가도 사라지고 강의도 머신러닝 필러닝을 알게 해서 가르치,
  "content_split": [
    "수업 운영이 주력구구식입니다.",
    "중요 과제 공지도 사흘 전에 기말 과제가 계획에 있다가도 사라지고 강의도 머신러닝 필러닝을 알게 해서 가르치는데 많은 내용을 그냥 알게 치고 지나",
    "본인이 따로 알아보고 찾아봐야 이해에 도움이 될 겁니다.",
    "다만 교수님은 중요십니다.",
    "무난합니다, 성적이 잘 주는 것 같아요.",
    "다만 실습이 조금 더 많았으면 좋겠어요."
  ],
  "content_normalized": [
    "수업 운영이 주력구구식입니다.",
    "중요 과제 공지도 사흘 전에 기말 과제가 계획에 있다가도 사라지고 강의도 머신러닝 필러닝을 알게 해서 가르치는데 많은 내용을 그냥 알게 치고 지나",
    "본인이 따로 알아보고 찾아봐야 이해에 도움이 될 겁니다.",
    "다만 교수님은 중요십니다.",
    "무난합니다, 성적이 잘 주는 것 같아요.",
    "다만 실습이 조금 더 많았으면 좋겠어요."
  ],
  "sentence_embeddings": [
    [
      -0.7870416641235352,
      0.7311447262763977,
      0.195688396692276,
      -0.03216171637177467,
      -0.10724299401044846,
      0.07256248593330383,
      0.9670011401176453,
      0.20776744186878204,
      -0.24033264815807343,
      0.12423437088727951,
      0.17025317251682281,
      0.7894136309623718,
      0.6610338091850281,
      -0.31083223232368622,
      0.06437689810991287,
      -0.16617177426815033,
      -0.492698073387146,
      0.02346283383667469,
      ...
    ]
  ]
}
  
```

* embedding vector 예시 (784차원)



ybigeee



모델링: chroma 생성 및 데이터 적재



수업 내용

- user3: 러시아에 대해 더 많이 알 수 있었던 것 같아요.

학점

- user2: 공부하면 성적 잘 주세요!!
- user3: 대면수업 때 집중만 잘 하시면 충분히 좋은 성적 받아가실 수 있습니다!

교수님
강의스타일
및 강의력

- user1: 교수님도 친절하시고
- user3: PPT 슬라이드가 1000장이 넘어가지만, 사진 위주의 필요없는 부분은 과감히 쳐내주셔서

시험 출제
스타일

- user3: 교수님께서 대면 수업 때 강조하시는 부분에서 대부분 문제들이 출제되기 때문에
- user4: 외울 건 많지만

ybigeee



모델링: Llama 3.1을 이용한 강의 요약

4

Llama 3.1 70B API



summarization



학정번호를 입력하세요 : EC02101

교수명을 입력하세요 : 김광환

[요약 결과]

- 강의명 : 거시경제학

- 강의 개요 : 이 강의는 국민생산, 소득, 실업, 인플레이션 등 거시경제학 주제에 대해 폭넓게 다룬다. 특히 New Keynesian 이론과 Real Business Cycle 이론을 비교해 보고, 개방화 시대를 맞이하여 개방경제의 거시경제모형을 중점적으로 다룬다. 또한 거시경제 이론이 한국경제의 상황에 어떻게 적용될 수 있는지 생각해본다.

- 강의평 :

- 학점 : 학점은 엄격하게 주어지며, A+를 받는 것은 어려운 편이다. 중간고사와 기말고사 문제들이 참고문제에서 나오므로, 교수님의 설명을 잘 이해하고, 참고문제를 풀어보면 좋은 성적을 받을 수 있다.

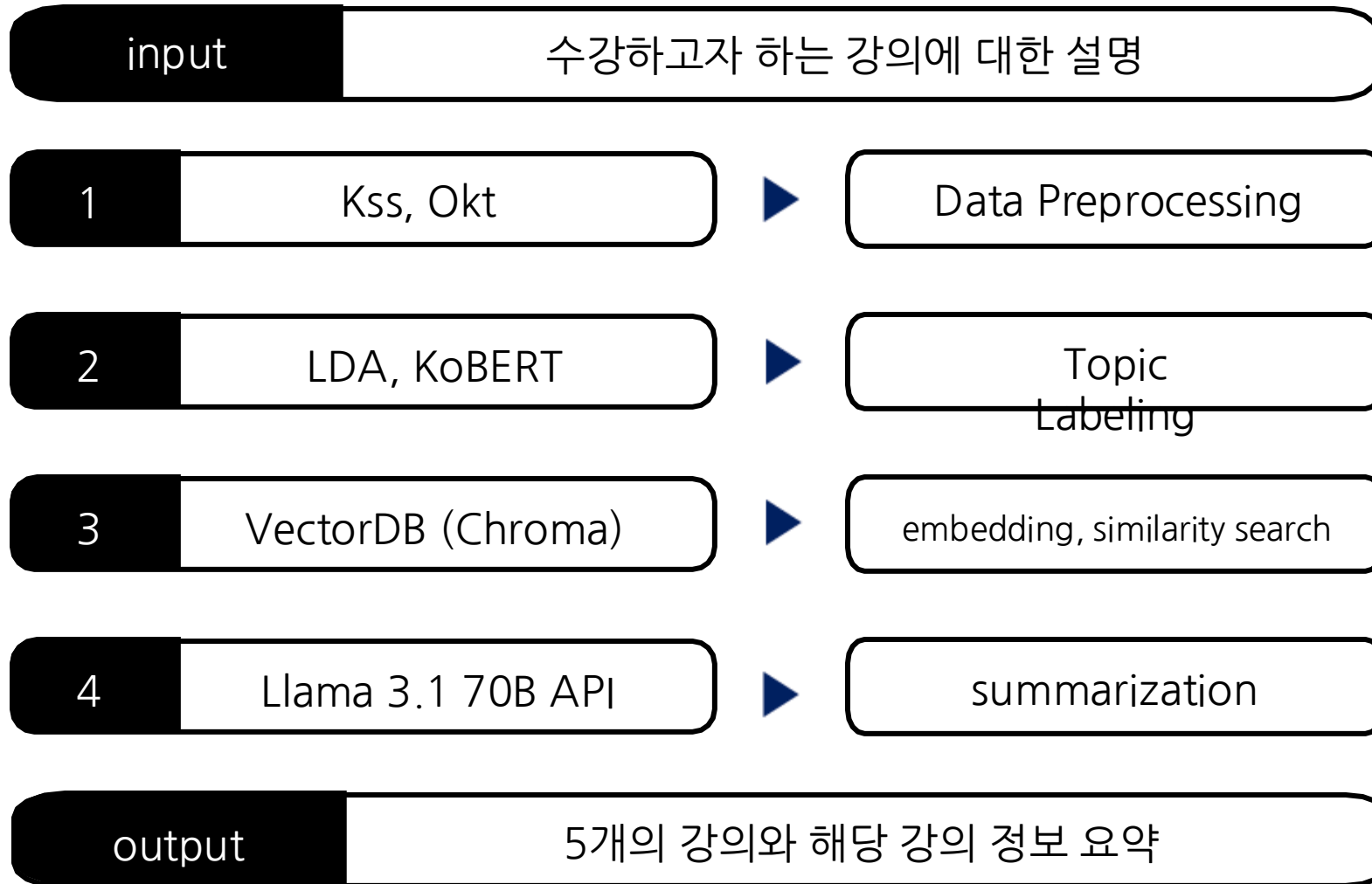
- 로드 : 로드는 적당한 편이나, 시험 문제들이 많고 길며, 틀리면 0점이 아니라 -1점이므로 주의해야 한다.

- 강의력 : 교수님의 영어 발음이 안 좋으며, 설명력이 떨어지지만, 그래프를 사용하여 설명하는 것은 좋다. 그러나 학생들과의 상호작용이 적으며, 질문을 잘 안 받아준다.

ybigeee



Pipeline





ybigeee



시연



ybigeee



한계점

1

1개의 문장 당 1개의 label만 할당 가능
→ 여러 개의 topic에 관한 문장일 시 classification 부정확

2

Topic Modeling 모델 성능 개선 필요 (validation set 기준 accuracy 0.64)

3

Llama 7.1 30B가 받을 수 있는 토큰 수 4000자로 제한
→ 4000자 초과 시 요약에 제외되는 자료 존재



ybigeee



Q & A