


# KLUE 강의 평가 요약 프로젝트

KUBIG 컨퍼런스 최종 발표: 클루섬

12기 이가영, 13기 금지현, 13기 기다연, 13기 박주영, 13기 홍석민

# Contents


- 
- 01. 프로젝트 주제
  - 02. 모델링
  - 03. API 개발
  - 04. 결과 소개
  - 05. Appendix

# 01. 프로젝트 주제

---

" 자연어 처리 기법을 사용하여  
교내 강의 평가 사이트 KLUE의 강의 별 텍스트 요약을 진행,  
보다 직관적으로 사용자들에게 유용한 정보를 제공 "

# 01. 프로젝트 주제

강의 만족도 

작성하신 리뷰를 확인하세요

작성하신 리뷰를 확인하세요

항목별 평가

학습량 5 난이도 5 강의력 5 성취감 5

2021.01.19 10시 11분에 남김

시험 난이도 자체는 무난했는데 주어진 분포함수 어떤 분포인지 초장에 찾아내지 못하면 모든 문제가 다 꼬입니다... 찾아냈다면 good. 그대로 쪽 한 문제씩 풀어나가면 되구요

만약 저처럼.. 찾아내지 못했을때(중간고사땀 찾았는데 기말고사에서는 못찾았습니다) 멘탈이 바사삭 깨질 수 있는데 이때 포기하시면 죽도 밥도 안됩니다. 간신히 남은 멘탈 부여잡고 포기하지 마시고! ☆☆☆문제 유형별로 풀이의 큰 틀은 정해져있으니 분포를 제외하고 쓸 수 있는건 다 쓰세요!!(MLE 구하는 방법 등등)☆☆ 그리고 수식이든 이론이든 구해볼 수 있는건 다 구해보면서 최대한 끄적이세요.

천사같은 조교님과 교수님께서 점수 줄 거리 최대한 찾아내서서 점수 주십니다. 기말 성적 확인할때 감사해서 오열할뻔했어요..


추가로) 지수분포와 감마분포, 카이제곱분포 연관관계 유의하면서 공부하시는걸 추천해요. 적분할때 Gamma function 적극 활용하면 꽤나 간단히 풀리는 문제들이 많습니다

역시 수리통계는 송교수님!

이 강의를 추천! 학점을 '기대 이상'(이)라 평가한 강의평입니다.

도움이 돼요 0

신고하기

강의 만족도 

작성하신 리뷰를 확인하세요

작성하신 리뷰를 확인하세요

항목별 평가

학습량 4 난이도 4 강의력 5 성취감 4

2021.01.18 05시 29분에 남김

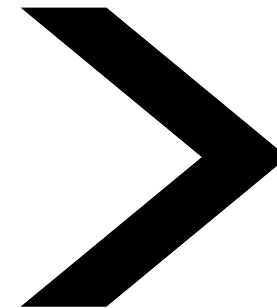
송성주 교수님께서 영어도 매우 잘하시고, 설명도 잘해주시고 강의력이 매우 좋습니다. 그래서 내용이 어려웠지만 잘 이해할 수 있었습니다. 중간 기말 모두 시험을 봤고, 과제가 매우 많은 편입니다.. 한 학기동안 8개 정도.. 그래도 수업 잘 듣고 필기 잘 했으면 하루 이틀 날 잡고 풀면 할 수 있습니다. 출석체크는 없었습니다.

학점은 엄청 잘 주시는 편은 아니었고 그러저럭 주셨습니다. 중간 미디언정도, 기말 Q3 넘고 과제 아주 조금 감점됐는데 B+ 받았어요.

학점을 '보통'(이)라 평가한 강의평입니다.

도움이 돼요 0

신고하기



" 교수님이 영어를 매우 잘하시고 난이도는  
꽤 있지만 성취감이 높습니다.  
학점을 엄청 잘 주시는 편은 아닙니다."  
- 수리통계학, 송성주 -

## 02. 모델링

---

- i. 요약 모델 스터디
- ii. 크롤링 및 데이터 전처리
- iii. Fine tuning
- iv. 모델 선정: KoBART

## 02. 모델링

### 요약 모델 스터디

#### 1. Attention Is All You Need

Attention 구조를 이용한 Transformation에 대한 이해

#### 3. BertSum: Text Summarization with Pretrained Encoders

extractive, abstractive model에 모두 사용 가능한 요약 모델

#### 2. Fine-tune BERT for Extractive Summarization

extractive summarization을 위한 BERT 변형 모델

#### 4. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension

임의로 noise를 주어 텍스트를 손상시킨 후 텍스트를 재구성하기 위한 모델을 학습하는 Transformer 기반 사전 학습 모델

요약 모델링을 위해 관련 논문 스터디를 병행

매주 공부한 내용을 돌아가면서 발표하며 모델에 대한 이해도 향상

논문	주제	진행날짜	발표자
Attention Is All You Need	model	2021년 9월 16일	기다현 박주영[학부재학 / 통계학과]
Fine-tune BERT for Extractive Summarization	model	2021년 9월 16일	Seokmin Hong 이가영[학부재학 / 통계학과]
BertSum: Text Summarization with Pretrained Encoders	model	2021년 9월 23일	Ji Hun Keom Seokmin Hong
BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension	model	2021년 9월 23일	박주영[학부재학 / 통계학과] 이가영[학부재학 / 통계학과]

## 02. 모델링

### 크롤링 및 데이터 전처리



# 02. 모델링

## Fine Tuning

강의 만족도 

항목별 평가 학습량 1 난이도 4 강의력 5 성취감 5

2021.08.04 04시 46분에 남김

andrewlau

계절을 들었는데 교수님이 진짜 설명을 이해하기 쉽게 잘 해주십니다. 중간에 과제나 숙제 이런것 하나도 없고 중간 기말 다 보통의 수준으로 문제가 제출되었습니다. 중간 기말 전에 교수님이 기출문제를 주시는데 진짜 완벽하게 풀 수 있을 정도로 공부할 많이 해야 합니다. 워낙 경제수학이다보니 기초가 약한 분들은 많이 어렵습니다.

학점을 '기대 이상'(이)라 평가한 강의평입니다.

도움이 됐요 1 신고하기

강의 만족도 

항목별 평가 학습량 5 난이도 5 강의력 1 성취감 1

2021.08.06 11시 24분에 남김

다리우스

당신 별 받을 거야!  
국강이지만 사실상 영강이 되어 버리는 매직...  
영어를 고집하지만 글씨체도, 발음도 알아 먹을 수가 없는 신비함.  
피드백 능력은 칭찬할만한 그.  
그렇지만 애초에 강의가 좋았다면 피드백을 장점으로 꼽을 일도 없었을 것.

학점을 '보통'(이)라 평가한 강의평입니다.

도움이 됐요 0 신고하기



```
import json
import pandas as pd

def read_json(path):
    DATA_DIR = '문서요약 텍스트_unzip'
    with open(DATA_DIR + path, 'r') as json_file:
        json_list = list(json_file)

    trains = []

    for json_str in json_list:
        line = json.loads(json_str)
        trains.append(line)
    df = pd.DataFrame(trains)
    return df

train_df = read_json('/1.Training/신문기사_1.train.json/train.json')
val_df = read_json('/2.Validation/신문기사_1.valId.json/valId.json')
train_df.head()
```

	media	id	article_original	abstractive	extractive
0	부산일보	360972161	[지난해 고령화와 유례가 드문 겨울 한파 등 영향으로 우리나라 사망자 수가 통계 작... 통계청이 발표한 '2018년 사망원인통계'를 보면 지난해 총 사망자 수는 관련 통계...		[4, 11, 18]
1	중도일보	356659913	[서산시의회(의장 임재관) 가중순.이수의 의원이 (사)한국지역신문협회에서 수여하는 ... 서산시 가중순 의원과 이수의 의원이 활발한 의정활동을 펼친 감사의 표시로 한국지역신...		[1, 3, 4]
2	무등일보	351718460	[지난 2004년 시작해 조선대 학생들의 대표적인 행사로 자리매김한 '조선대 국토대... '조선대의 새로운 비상을 꿈꾸다'를 슬로건으로 진행되어 단체생활을 통해 협동심과 ...		[0, 2, 4]
3	이데일리	335868123	[서울시는 신학기가 시작되는 다음달 4월부터 고등학교 3학년 무상급식을 실시한다고 ... 서울시가 다음달 4월부터 서울 시내 319개 고등학교 3학년 8만4700명을 대상으...		[0, 1, 2]
4	서울신문	351443347	[미국인 선교사가 우간다에서 의사 병서를 하며 의료 시설을 운영한 혐의로 지역 시민... 미국인 선교사가 우간다에서 의사 행세를 하며 두 아이의 죽음과 관련돼 있다며 지역 ...		[0, 1, 2]

KLUE 리뷰의 텍스트와 가장 유사하고 평서문 기반의 Abstractive Summarization Label을 제공하는 데이터 선택



## 02. 모델링

## 모델 선정: KoBART, KoT5, BertShared

약 2-3주 동안 3 팀으로 나누어 KoBART, KoT5, BertShared 모델을 학습

세 모델의 학습 결과, 가장 그럴듯한 요약해낸 KoBART를 최종 모델로 선정

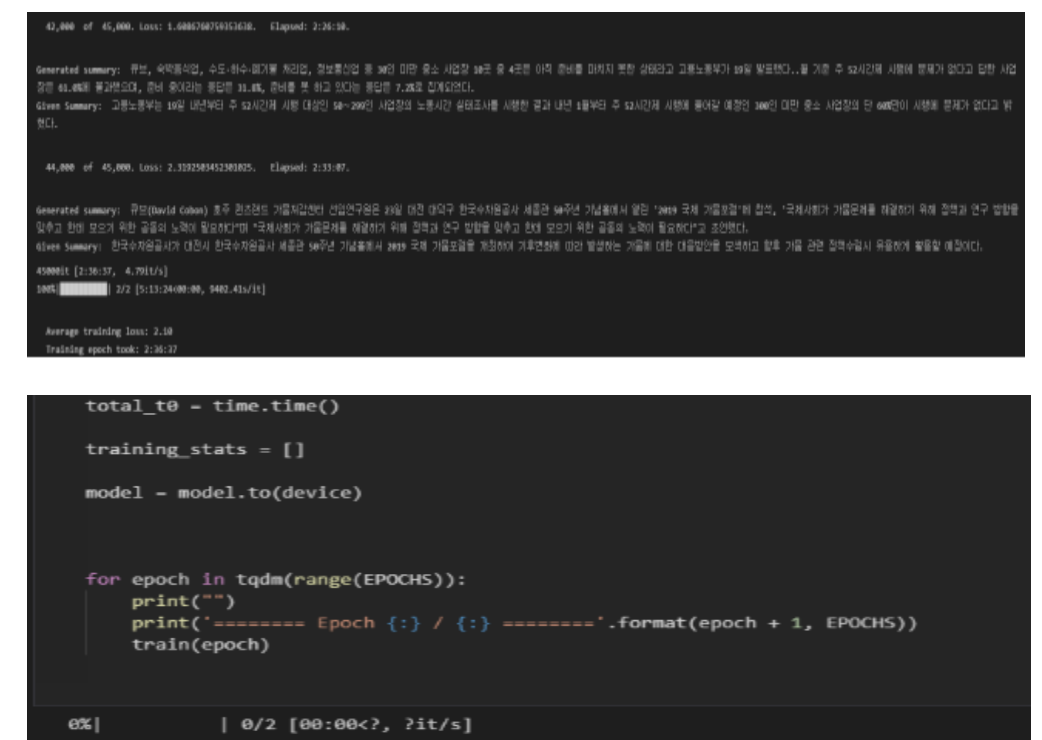
## KoBART



# BertShared



# T5



## 02. 모델링

### KoBART로 진행한 강의평 요약

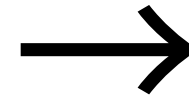
'ππ 매우 어려웠어요... 과제, 공부하는 데에도 시간 많이 쓰고 그랬는데 시험에서 정작 힘을 못 썼어요. 그래도 별 후회도 안했습니다. 중간고사까지는 확률론 열심히 공부한 게 도움이 될 텐데 기말은 비슷합니다. 교수님 수업 잘하시는데 내용 이해하는 것은 개인 몫입니다. 시간 정말 많이 쓰셔야 해요.'

'배우는 내용자체가 많아서 어렵다기보다는, 개념자체가 어렵습니다. 솔직히 정확히 이해하는 사람 몇 없는 것 같습니다. 그냥 문제풀때 이 개념을 어떻게 쓰는가정도만 알고, 여러번 반복해서 풀다보니 그냥 그렇게 풀어야하는가보다 하고 풀었습니다. 근데 시험문제는 개념에 관한 것이 아니라, 그냥 문제를 최대한 풀기 힘들게 만들어 놓은 형태입니다.

⋮

'교수님께서 진짜 잘 가르치신다 영어도 되게 잘하시고 강의력도 좋으시고 친절하시다 중간때까지는 그렇게 어렵진 않은데 기말부분은 수업때 조금 못 알아들을수도 있다 과제가 엄청 자주 나오고 공부해야될게 많다 수업은 재밌고 정말 많이 배우는데 학습량이 너무 많아서 조금 부담되기는한다 그래도 수통 어차피 들어야되면 이분께 듣는걸 추천'

송성주 교수님의 수리통계학 강의평을  
KoBART를 이용해 요약한 결과



**"통계수학에서 다루는 미분과 적분과 확률론에서 다루는 mgf, iterative expectation 등 중요한 내용에 대한 숙지가 부족한 경우는 성적이 매우 안좋을 수 있으며 통계학과의 과목 특성상 이 내용이 저 내용과 연결이 되는 그 밀도가 다른 과목보다 월등이 높기 때문에 내용 이해는 물론 시험에 대한 자세도 중요합니다."**

# 03. API 개발

---

i. FLASK API 스터디

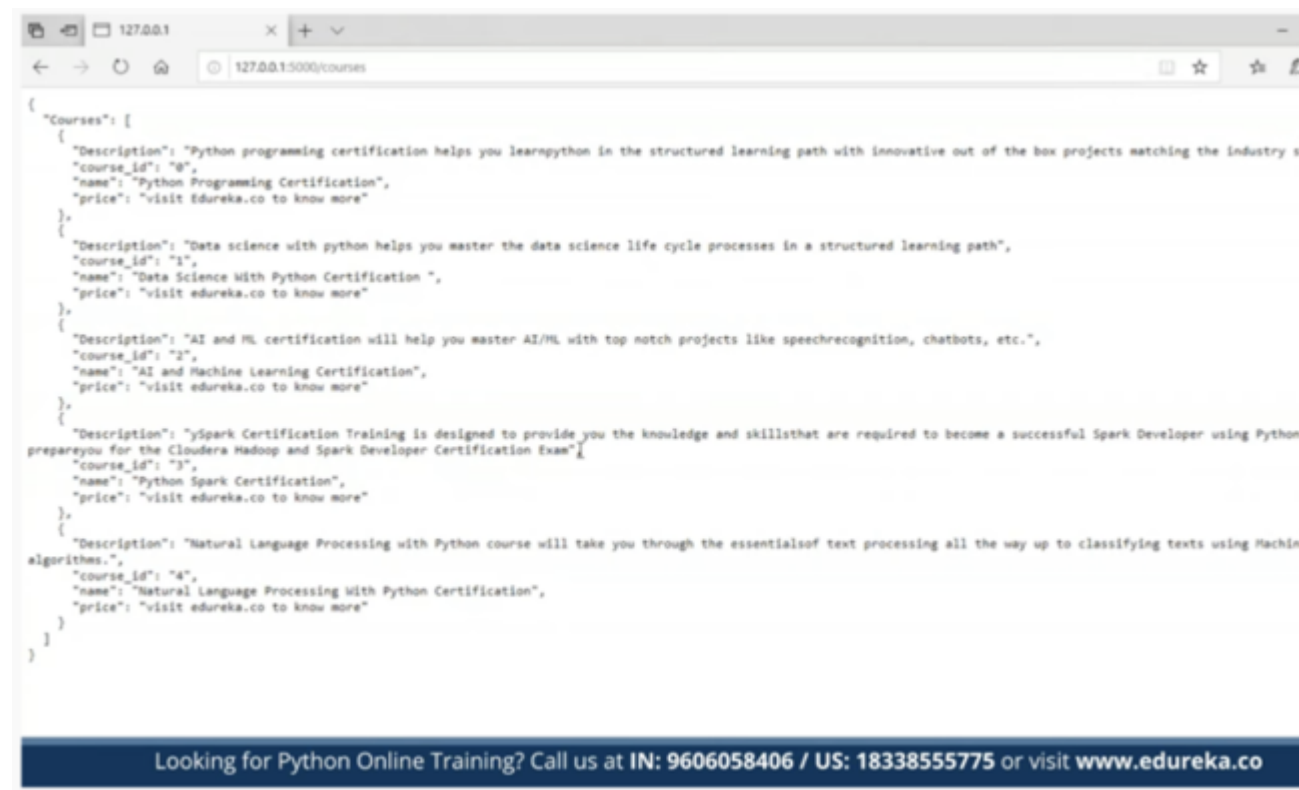
ii. REST API 구현

# 03. API 개발

## FLASK API 스터디

### 2주에 걸쳐 API 스터디 진행

- > 가볍고 심플한 특징을 갖는 프레임워크인 FLASK API를 선택
- > YouTube 채널 "Edureka" 통해 REST API 기반의 모델 배포 학습
- > 그 외 다양한 학습자료 바탕으로 FLASK 작동원리 이해



FLASK 기본 메소드부터 작동원리 이해

### HTTP methods

Sl.no	Method	Description
1	GET	Sends data in unencrypted form to server
2	HEAD	Same as GET, but without response body
3	POST	Used to send HTML form data to server.
4	PUT	Replaces all current representations of target resource with uploaded content
5	DELETE	Removes all current representations of target resource given by URL

예) login.html으로 아래 HTML 파일 저장

```
<html>
<body>
  <form action = "http://localhost:5000/login" method="post">
    <p>Enter name:</p>
    <p><input type = "text" name="nm"/></p>
    <p><input type = "submit" value="submit"/></p>
  </form>
</body>
</html>
```

```
from flask import Flask, redirect, url_for, request
app = Flask(__name__)

@app.route('/success/<name>')
def success(name):
    return 'Welcome %s' % name

# login.html과 연결해서 POST/GET method 중에 정해서 사용
@app.route('/login', methods=['POST', 'GET'])
def login():
```

YouTube 채널 "Edureka" 통해 REST API 기반 모델 배포 학습

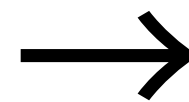
# 03. API 개발

## REST API 구현

### REST API

- > 웹의 장점을 최대한 활용할 수 있는 아키텍처
- > 행위에 대한 정의는 HTTP METHOD를 통해 하는 것이 REST한 API를 설계하는 중심 규칙

```
16 @app.route("/", methods=["GET", "POST"])
17 def index():
18     if request.method == "GET":
19         return render_template("index.html")
20     if request.method == "POST":
21
22         # data = {"강의명": "", "요약문": ""}
23
24         if request.form['items1'] == "org_text1":
25             text = str(
26                 "ㅠㅠ 매우 어려웠어요... 과제, 공부하는 데에도 시간 많이 쓰고 그랬는데 시험에서 정작 힘을 못 썼어요.
27             lecture = "수리통계학"
28 > elif request.form['items1'] == "org_text2": ...
32 > elif request.form['items1'] == "org_text3": ...
```



웹에서 요약 요청을 받으면 모델을 통해 요약문(리소스)을 생성하고,  
Json 형식으로 반환

METHOD	역할
POST	POST를 통해 해당 URI를 요청하면 리소스를 생성합니다.
GET	GET를 통해 해당 리소스를 조회합니다. 리소스를 조회하고 해당 도큐먼트에 대한 자세한 정보를 가져온다.
PUT	PUT를 통해 해당 리소스를 수정합니다.
DELETE	DELETE를 통해 리소스를 삭제합니다.





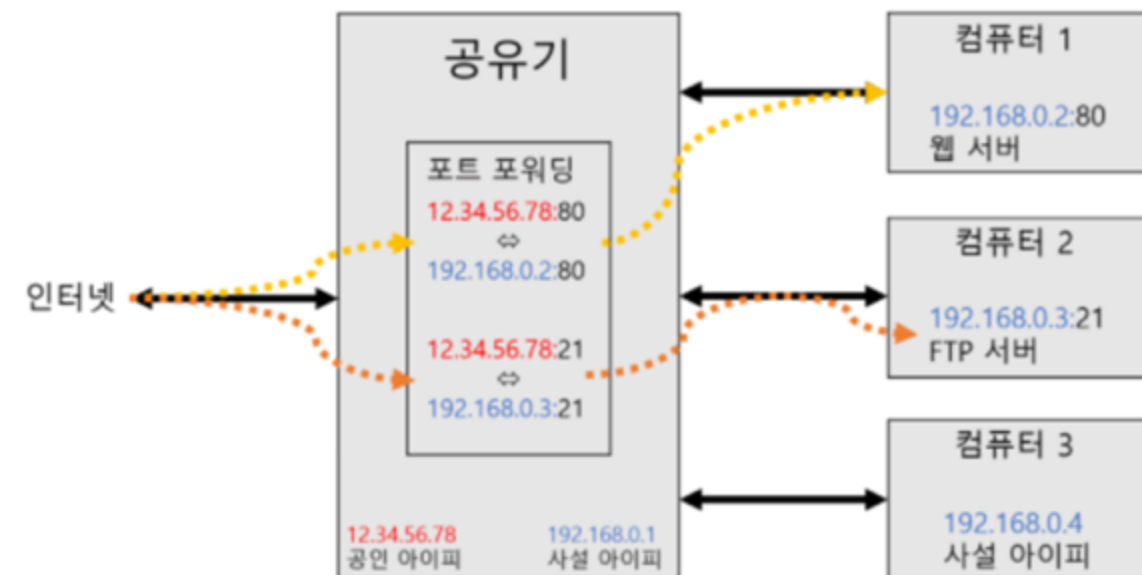
# 03. API 개발

## 모델 배포

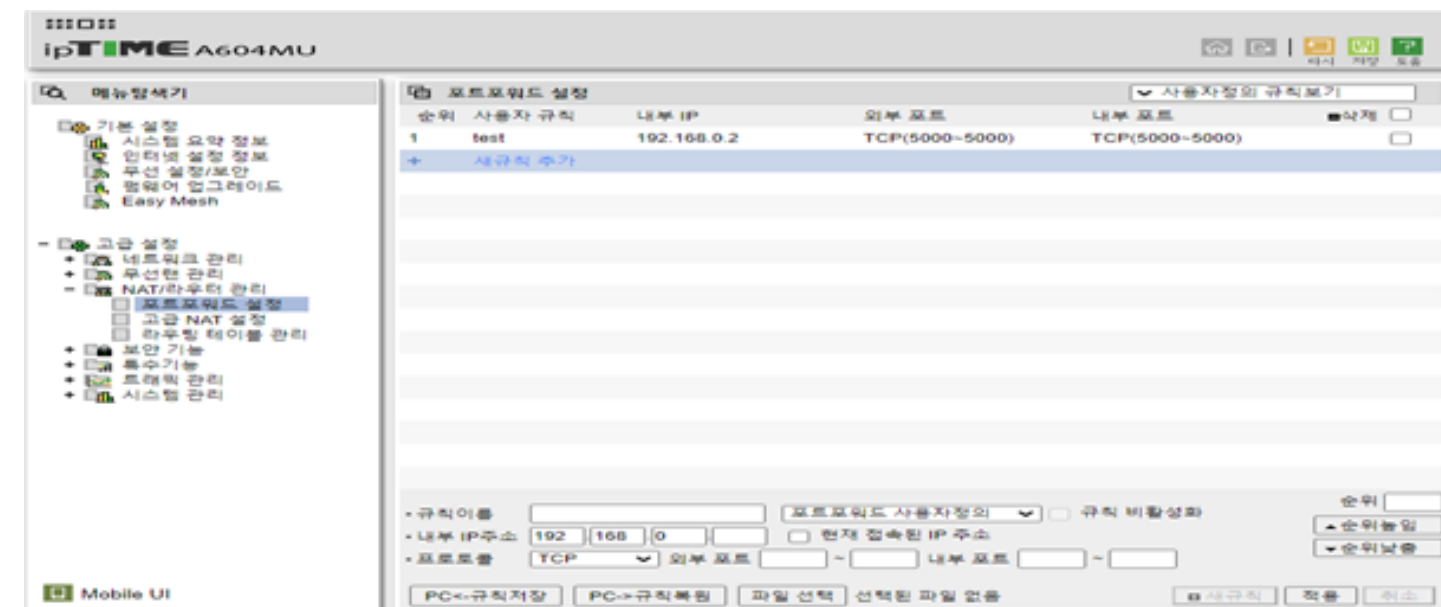
### 포드포워딩

- > 개인 PC의 API 코드를 외부에서 사용할 수 있게 배포 필요
- > 포드포워딩을 통해 KLUUE 개발자들이 인터넷을 통해 요약 API에 접근 가능
- > 외부 인터넷에서 공유기로 접근하는 것을 포드포워딩을 통해 개인 PC 포트에 연결

포드포워딩을 통한 API 접근



포드포워딩을 통한 개인 PC 포트 연결



## 04. 결과 소개

---

- i. 그러나...
- ii. API 시연 영상

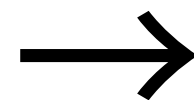
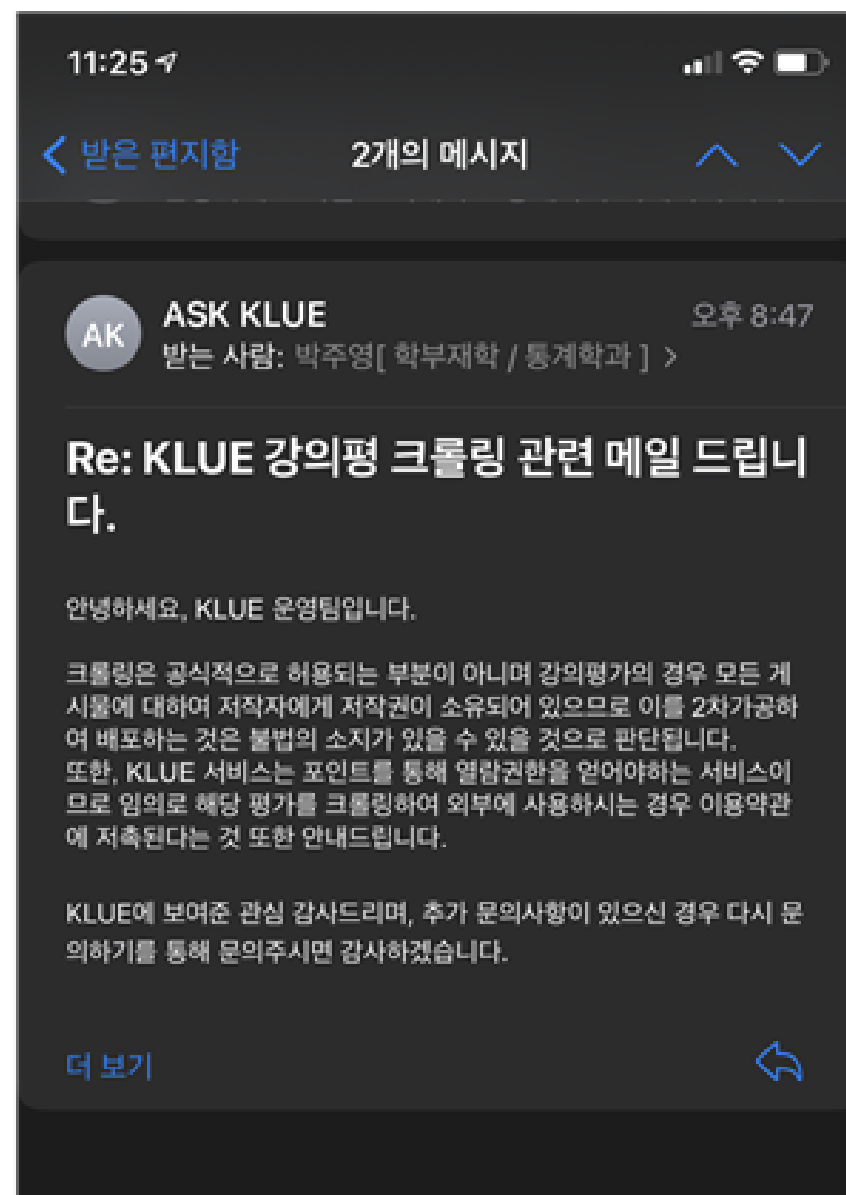


## 04. 결과 소개

그러나...

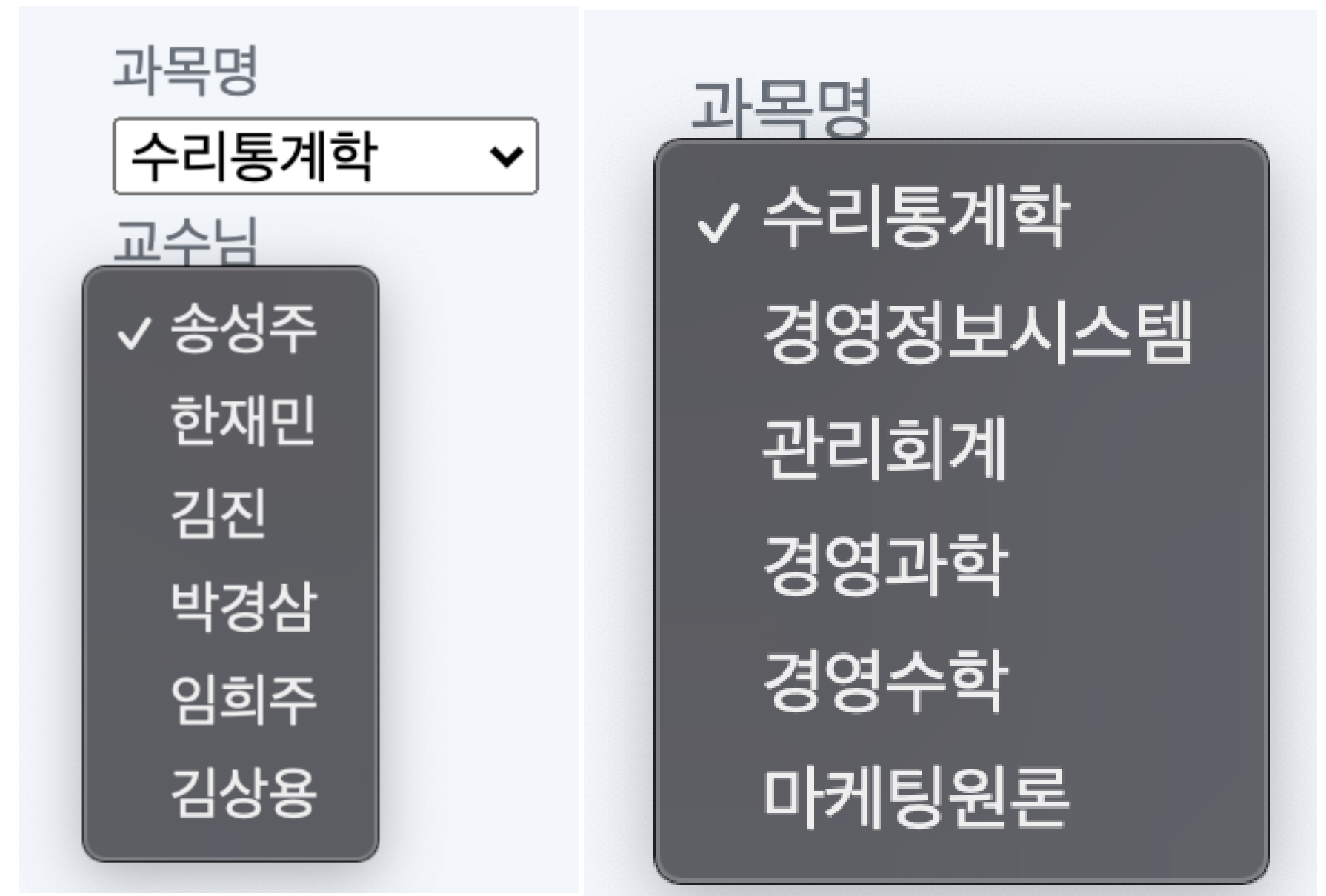
### 기존 계획

- > KLUE 측에 요청해서 데이터 제공 받기
- > KLUE 리뷰 크롤링하기



### 수정 방향

- > 6개 강의에 대해 각각 10개의 리뷰만 샘플로 수집
- > 10개 리뷰를 바탕으로 각 강의에 대한 요약 성능 확인 및 API 구현



# 04. 결과 소개

API 시연 영상

클루 강의평 요약 서비스  
KUBIG 21-2 장기 프로젝트

과목명  
수리통계학 ▼  
교수님  
송성주 ▼

요약하기

요약 결과

# 05. Appendix

관련 코드 및 스터디 내용

KLUE 프로젝트 노션

<https://www.notion.so/keom/KLUE-2333d995ff16488d8fe6320ef2a80aec>

KLUE 프로젝트 티스토리

<https://kubig-2021-2.tistory.com/>