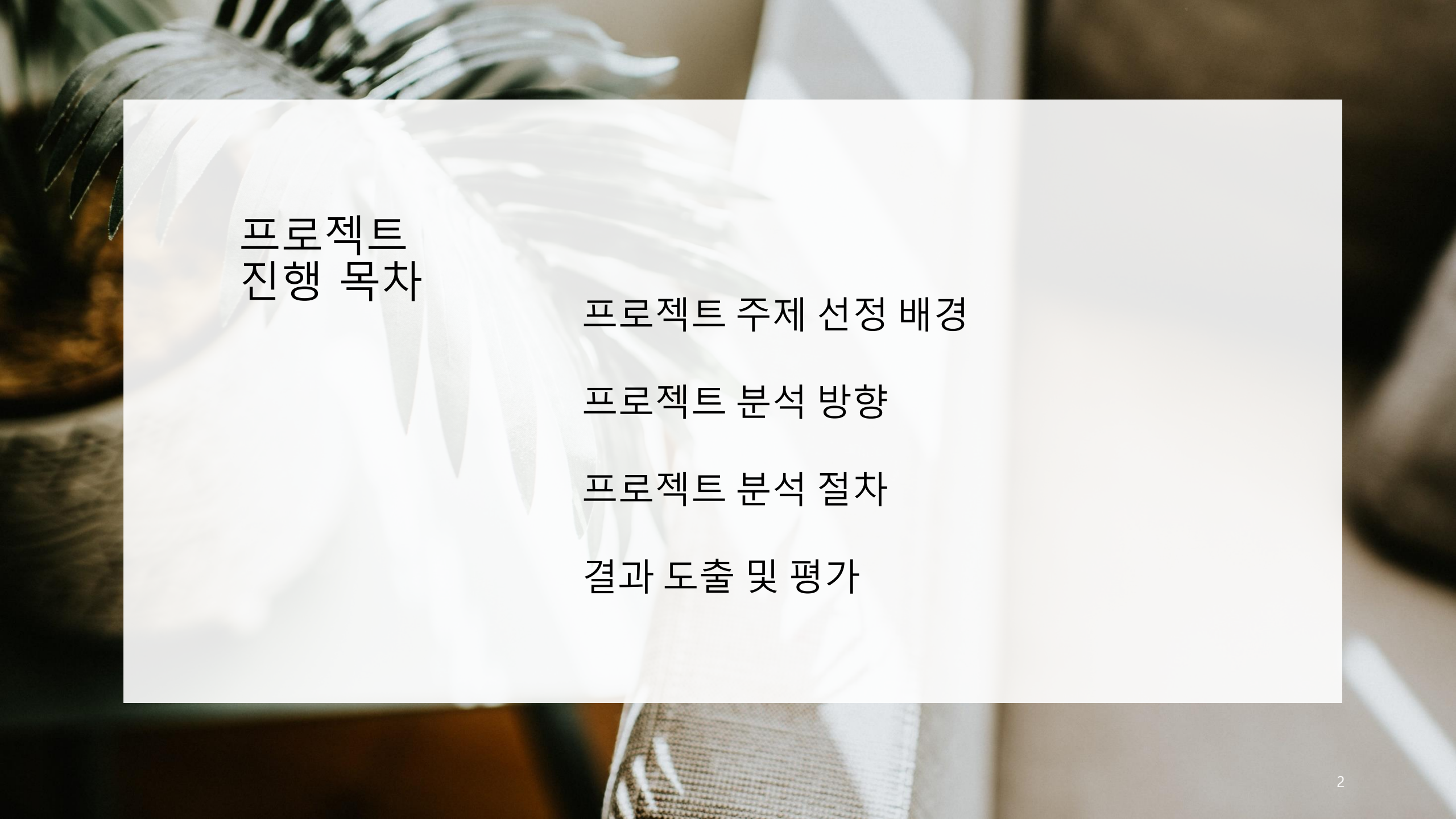




대학 교과목 분석을 통한 학과 추천 모델 연구

- 1 조 -



프로젝트 진행 목차

프로젝트 주제 선정 배경

프로젝트 분석 방향

프로젝트 분석 절차

결과 도출 및 평가

프로젝트 주제 선정 배경

대학 입시요강도 중요하지만 대학 재학중에 이수하게 될 학과별 커리큘럼에 대한 정보를 토대로 "실제로 어떤 교과목들이 자신이 원하는 대학 학과와 연관되어 있을까?" 에 대한 의문에서 해당 주제를 선정하게 되었다.



기대 수요층

- 대학 입학 준비 고교생, 재수생
- 고등학교 3학년 담임 선생님



기대 효과

- 대입 준비 과정에서 학생들이 배우고 싶은 과정과 연관성이 있는 학과들을 추천한다.
- 새롭게 개편된 학과의 수업, 신설 학과의 수업들로 기존 학과와 유사한 학과를 확인할 수 있다.



프로젝트 분석 방향

1. 학과별 교과목 커리큘럼 분석

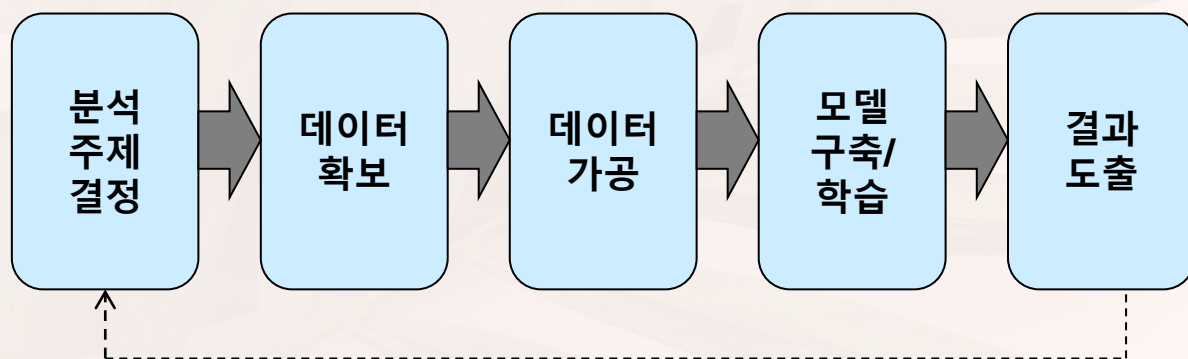
대학에 개설된 학과별 이수 교과목, 커리큘럼 데이터를 수집, 분석.

2. 최적 학과 Proposal

분석된 전공 교과목들을 토대로 가장 연관성이 있는 최적화된 학과를 추천 할 수 있다.



프로젝트 분석 절차



작업 환경 및 분석 도구

1. 작업 환경 : 코랩 구글

2. 분석도구 및 기술

- Python, Keras
- KoNLPy
- TensorFlow
- Word Embedding
- DNN



프로젝트 분석 절차 – 데이터 확보

“ 대학 학과별 개설 교과목 (커리큘럼) 데이터 수집 “

WORKNET | 일자리 검색 | 청년 | 여성 | 중장년

홈 > 직업·진로 > 학과정보 > 학과 검색

직업·진로 | **학과 검색**

직업심리검사 소개
직업심리검사 실시
청소년 심리검사 실시
성인용 심리검사 실시
검사결과 보기
검사결과 상담
검사결과 상담 FAQ

한국직업정보
직무기반 직업추천
한국직업전망
한국직업사관
한국직업사관
직업 동영상 VR
직업인 인터뷰
다양한 직업세계
미래직업 탐구

학과정보
학과 검색
관공 진로가이드
학과정보 FAQ
학과정보 동영상

진로상담
상담신청
상담 FAQ

자료실
직업진로정보서
연구보고서

취업 동영상
청년층
경력단절여성
중장년층

학과 검색

국어·국문학과
인문계열 > 국어·국문학과

요약보기 | **학과소개** | 개설대학 | 모집현황 | 진출가능직업

학과 소개
학과의 특성을 개괄적으로 소개합니다.
또한 학과 공부에 필요한 핵심이나 흥미, 전공과 관련하여 취득할 수 있는 국가(기술) 자격 및 면허를 소개합니다.

· 개요
국어국문학과에서는 유문론, 문법론, 어휘론 등의 '국어학'과 고전문학 및 현대문학을 공부하는 '국문학'분야를 공부합니다. 국어국문학과에서는 언어와 문학 작품에 대한 포괄적인 이론적 지식을 알고 우리말과 글이 어떠한 문법의 구조를 가지고 어떻게 변천해 왔는지, 사회변화에 따라 글이 어떻게 영향을 받았는지에 대해서도 공부합니다. 또한 문학작품에 대한 분석을 통해 우리 문학의 본질과 우리 국민들의 정서를 파악합니다. 학부과정에서는 '국어학'과 '국문학'에 대한 전반적인 교과목을 모두 배우고 대학원에 진학한 후 세부전공으로 나누어지는 경우가 많습니다.

· 특성 및 흥미
국어국문학과는 인문과학의 층위가 되는 학과로, 무엇보다 언어와 문학에 관심과 소질이 있는 것이 좋습니다. 언어와 문학은 사회적 환경에 영향을 받기 마련이므로 사회변화를 읽을 줄 알면 더욱 좋습니다. 또한 언어, 영화, 미술, 광고 등 다양한 문화예술 장르에 흥미를 가지고 있으면 문학작품을 이해하고 창작하는데 도움이 됩니다. 상상력과 창의력 그리고 글쓰기, 읽기에 대한 흥미가 필요합니다. 또한 한국어 문법과 음운 규칙 등 국어학에 대한 공부를 위해 호기심과 분석력도 필요합니다.

· 관련학과
국어국문학과, 한국어문학과, 한국어학과, 한문학과, 미디어콘텐츠전공, 스토리텔링학과, 통상언어전공, 한국어교원학과, 한국어통번역전공

· 교과목
* 국어학 과목 : 국어학개론, 국어사, 국어방언학, 국어의미론, 국어문법론, 국어음운론, 국어학 연구, 한국어 형태론, 한국어 음성학, 한국어 어휘론, 한국어 정책론, 한국어정보처리, 한국어발달사, 응용한국어학 등
* 국문학 과목 : 국문학개론, 한국현대문학사, 한국현대작가론, 한국고전문학사, 고전작품강독, 문학작품과 한국사, 현대시론, 현대비평론, 현대회고론, 한국근대문학의 형성, 한국한문학개론, 구비문학론, 근세시가론, 고전시가론, 고전소설론, 현대소설론 등

· 취득자격면허
국가자격 : 박물관및미술관준학예사
민간자격 : 논술지도사, 독서지도사, 아동독서지도사, 시니어독서지도사, 독서코칭전문가, 독서논술지도사 등

· 관련학과

국어국문학과, 한국어문학과, 한국어학과, 한문학과, 미디어콘텐츠전공, 스토리텔링학과, 통상언어전공, 한국어교원학과, 한국어통번역전공

· 교과목

* 국어학 과목 : 국어학개론, 국어사, 국어방언학, 국어의미론, 국어문법론, 국어음운론, 국어학 연구, 한국어 형태론, 한국어 음성학, 한국어 어휘론, 한국어 정책론, 한국어정보처리, 한국어발달사, 응용한국어학 등

* 국문학 과목 : 국문학개론, 한국현대문학사, 한국현대작가론, 한국고전문학사, 고전작품강독, 문학작품과 한국사, 현대시론, 현대비평론, 현대회고론, 한국근대문학의 형성, 한국한문학개론, 구비문학론, 근세시가론, 고전시가론, 고전소설론, 현대소설론 등

· 취득자격면허

국가자격 : 박물관및미술관준학예사

민간자격 : 논술지도사, 독서지도사, 아동독서지도사, 시니어독서지도사, 독서코칭전문가, 독서논술지도사 등

워크넷에서 운영하는 학과별 이수과목 데이터 이용

프로젝트 분석 절차 – 데이터 확보

“ 대학 학과별 개설 교과목 (커리큘럼) 데이터 수집 “



2021 학과(전공) 분류 자료집

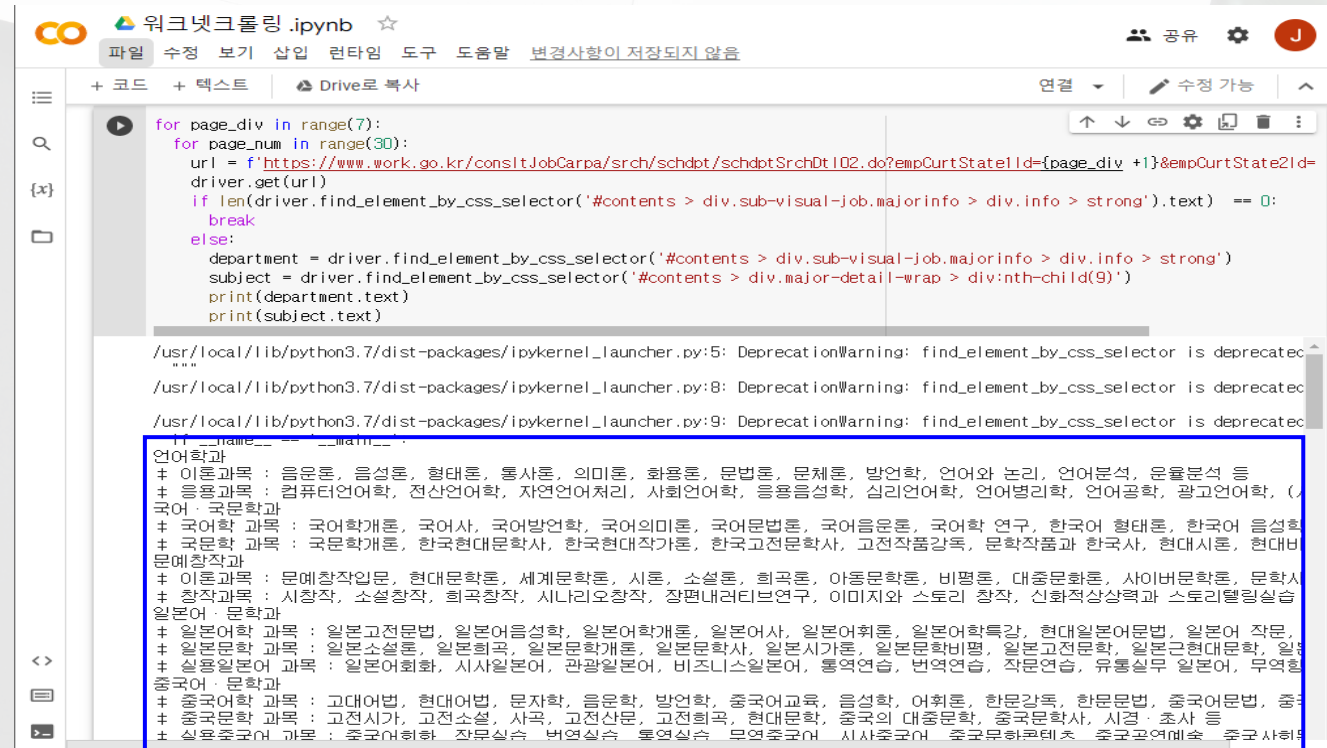
대분류	중분류	소분류	학과명	학과코드
교육계열 03 (계속)	교육일반 0301	교육학 030101	평생교육진흥학과	U03010100077
			평생교육학 - HRD연계전공	U03010100078
	특수교육 0303	특수교육학 030301	특수교육(초등특수교육)	U03030100047
			특수교육(중등특수교육)	U03030100048
		특수교육(유아특수교육)	특수교육(유아특수교육)	U03030100049
	중등교육 0305	영어교육 030501	글로벌영어교육학과	U03050100040
	공학계열 04	건축 0401	건축 - 도시계획학과	U04010200115
			스마트시티건축학전공	U04010200116
건축학과(4년제)			U04010200117	
해양건축 - 에너지자원공학부 (공간디자인전공, 건축정책공학전공)			U04010200118	
토목 · 도시 0402		스마트시티건축공학과	U04010200119	
		사회환경공학과	U04020100202	
		스마트건축토목공학부	U04020100203	
		건설시스템안전공학과	U04020100204	
도시공학 040202		토목안전공학과	U04020100205	
		물류 · 환경 · 도시인프라공학부(건설공학전공)	U04020100206	
		도시융합시스템공학과	U04020200058	
		스마트시티공학전공	U04020200059	
		그린스마트시티학과	U04020200060	
		스마트도시학부	U04020200061	
비공 · 공중 0403		차량공학과 040301	스마트도시건설전공	U04020200062
				U04030100037
		항공학 040302	항공기계공학전공	U04030100038
			항공기계공학전공	U04030200079
	항공운송학과 040303	스마트드론공학과	U04030200080	
		항공운송학과	U04030200081	
공학계열 04	해안공학 040303	항공우주시스템공학부(항공공학전공)	항공우주시스템공학부(항공공학전공)	U04030200082
			항공우주시스템공학부(항공공학전공)	U04030200083
		조선 · 해양개발공학부(조선공학전공)	조선 · 해양개발공학부(조선공학전공)	U04030300136
			조선 · 해양개발공학부(조선공학전공)	U04030300137
		조선 · 해양개발공학부(조선공학전공)	조선 · 해양개발공학부(조선공학전공)	U04030300138
			조선 · 해양개발공학부(조선공학전공)	U04030300139
	해안공학 040303	조선 · 해양개발공학부(조선공학전공)	조선 · 해양개발공학부(조선공학전공)	U04030300140
			조선 · 해양개발공학부(조선공학전공)	U04030300140
		조선 · 해양개발공학부(조선공학전공)	조선 · 해양개발공학부(조선공학전공)	U04030300140
			조선 · 해양개발공학부(조선공학전공)	U04030300140

공학계열 04	건축 0401	건축학 040102	건축 · 도시계획학과
			스마트시티건축학전공
			건축학과(4년제)
			해양건축 · 에너지자원공학부 (공간디자인전공, 건축정책공학전공)
	토목 · 도시 0402	토목공학 040201	스마트시티건축공학과
			사회환경공학과
			스마트건축토목공학부
			건설시스템안전공학과
		도시공학 040202	토목안전공학과
			물류 · 환경 · 도시인프라공학부(건설공학전공)
		도시공학 040202	도시융합시스템공학과
			스마트시티공학전공
		도시공학 040202	그린스마트시티학과
			스마트도시학부

한국교육개발원 2021 학과(전공) 분류 자료집 이용

프로젝트 분석 절차 – 데이터 확보

“대학 학과별 개설 교과목
(커리큘럼)을 웹 크롤링 진행 “



```
for page_div in range(7):
    for page_num in range(30):
        url = f'https://www.work.go.kr/consil/JobCarpa/srch/schdpt/schdptSrcht102.do?empCurtState1Id={page_div + 1}&empCurtState2Id={page_num + 1}'
        driver.get(url)
        if len(driver.find_element_by_css_selector('#contents > div.sub-visual-job.majorinfo > div.info > strong').text) == 0:
            break
        else:
            department = driver.find_element_by_css_selector('#contents > div.sub-visual-job.majorinfo > div.info > strong')
            subject = driver.find_element_by_css_selector('#contents > div.major-detail-wrap > div:nth-child(9)')
            print(department.text)
            print(subject.text)
```

...
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:5: DeprecationWarning: find_element_by_css_selector is deprecated
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:8: DeprecationWarning: find_element_by_css_selector is deprecated
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:9: DeprecationWarning: find_element_by_css_selector is deprecated

언어학과
+ 이종과목 : 음운론, 음성론, 형태론, 통사론, 의미론, 화용론, 문법론, 문체론, 방언학, 언어와 논리, 언어분석, 문물분석 등
+ 응용과목 : 컴퓨터언어학, 전산언어학, 자연언어처리, 사회언어학, 응용음성학, 심리언어학, 언어병리학, 언어공학, 광고언어학, (, 국어, 국문학과
+ 국어학 과목 : 국어학개론, 국어사, 국어방언학, 국어의미론, 국어문법론, 국어음운론, 국어학 연구, 한국어 형태론, 한국어 음성학
+ 국문학 과목 : 국문학개론, 한국현대문학사, 한국현대작가론, 한국고전문학사, 고전작품강독, 문학작품과 한국사, 현대시론, 현대비문예창작과
+ 이종과목 : 문예창작입문, 현대문학론, 세계문학론, 시론, 소설론, 희곡론, 아동문학론, 비평론, 대중문화론, 사이버문학론, 문학사
+ 창작과목 : 시창작, 소설창작, 희곡창작, 시나리오창작, 장편대러티브연구, 이미지와 스토리 창작, 신화적상상력과 스토리텔링실습
일본어 : 문학과
+ 일본어학 과목 : 일본고전문법, 일본어음성학, 일본어학개론, 일본어사, 일본어회화, 일본어학특강, 현대일본어문법, 일본어 작문,
+ 일본문학 과목 : 일본소설론, 일본희곡, 일본문학개론, 일본문학사, 일본시가론, 일본문화비평, 일본고전문학, 일본근현대문학, 일
+ 실용일본어 과목 : 일본어회화, 시사일본어, 관광일본어, 비즈니스일본어, 통역연습, 번역연습, 작문연습, 유통실무 일본어, 무역학
중국어 : 문학과
+ 중국어학 과목 : 고대어법, 현대어법, 문자학, 음운학, 방언학, 중국어교육, 음성학, 어휘론, 한문강독, 한문문법, 중국어문법, 중
+ 중국문학 과목 : 고전시가, 고전소설, 사곡, 고전산문, 고전희곡, 현대문학, 중국의 대중문학, 중국문학사, 시경·초사 등
+ 실용중국어 과목 : 중국어회화, 작문실습, 번역실습, 통역실습, 무역중국어, 시사중국어, 중국문화콘텐츠, 중국공연예술, 중국사회

프로젝트 분석 절차 – 데이터 확보

“웹 크롤링 후 커리큘럼 데이터 정제 “

- 엑셀로 변환
- 특수문자 처리
- 외국어, 숫자 처리
- csv 파일로 저장
- 5980쌍 제작

학과	전공수업
건축학과	건축학개론, 건축이론, 건축사, 건축설계, 건축구조의 이해, 주택계획, 건축구조, 건축재료, 건축설비, 건축법규, 건축캐드, ;
건축설비공학과	건축개론, 건축제도, 동역학, 구조역학, 재료역학, 건축재료공학, 건축구조, 건축시공, 건축설비, 건설기술, 건설경영, 건축;
조경학과	조경학원론, 조경사, 조토양학, 지형학, 서양조경문화사, 조경수목학, 조경계획방법론, 조경구조공학, 조경적산학, 조경소재
토목공학과	공학수학, 수리학, 토질역학, 구조공학, 철근콘크리트공학, 토목시공학, 토목설계, 방재학개론, 재료역학, 토질역학, 지반회
도시공학과	도시계획, 도시발달사, 도시개발, 도시설계, 교통계획, 교통공학, 지역계획, 환경계획, 도시학개론, 도시디자인, 측량학, 토;
지상교통공학과	교통류이론, 교통안전공학, 교통수요분석, 교통계획, 도로설계, 교통학개론, 계량분석, 공학수학, 교통공학론, 교통통계학,
항공학과	항공무기체계, 헬리콥터정비, 항공열역학, 항공기구조물 수리실습, 가스터빈엔진일반, 장탈작동, 항공기재료일반, 공구실;
해양공학과	선박유체역학, 조선해양공학계획, 기초구조동역학, 용접구조설계, 물리학, 해양플랜트개론, 전산선박제도, 공업역학, 재료
기계공학과	고체역학, 열역학, 동역학, 유체역학, 재료역학, 기계설계, 기계공학실험, 공학물리, 수학, 공학전산응용, 에너지공학, 열전;
재료공학공학과	재료설계학, 재료선택 및 활용, 재료소성론, 재료기기분석, 재료설계학, 반도체공학, 디스플레이공학, 상평형, 나노재료, 소
자동차공학과	열역학, 동역학, 유체역학, 재료역학, 자동차재료, 전기전자공학, 자동차공학개론, 자동차구조실습, 공학수학, 정역학, 고차
전기공학과	전기회로, 회로이론, 전자기학, 자동제어, 인공지능제어, 배전계통운용, 신호 및 시스템, 디지털회로, 제어공학, 전자기장, ;

계열	학과	전공수업
공학계열	건축학과	건축학개론
공학계열	건축학과	건축이론
공학계열	건축학과	건축사
공학계열	건축학과	건축설계
공학계열	건축학과	건축구조의 이해
공학계열	건축학과	주택계획
공학계열	건축학과	건축구조
공학계열	건축학과	건축재료
공학계열	건축학과	건축설비
공학계열	건축학과	건축법규
공학계열	건축학과	건축캐드
공학계열	건축학과	건축프로그래밍
공학계열	건축학과	도시개발
공학계열	건축학과	조경설계
공학계열	건축학과	생태건축
공학계열	건축학과	친환경건축
공학계열	건축학과	인테리어계획론

프로젝트 분석 절차 – 데이터 가공

“ 데이터 가공, 전처리 ”

- mecab 으로 토큰화 진행
 - 토큰화 처리: 텍스트 길이가 긴 경우 문장과 단어를 구분하는 방식

```
[ ] !pip install konlpy  
  
!bash <(curl -s https://raw.githubusercontent.com/konlpy/konlpy/master/scripts/mecab.sh)
```

```
import os  
os.chdir('/tmp/')  
!curl -LO https://bitbucket.org/eunjeon/mecab-ko/downloads/mecab-0.996-ko-0.9.1.tar.gz  
!tar xzfv mecab-0.996-ko-0.9.1.tar.gz  
os.chdir('/tmp/mecab-0.996-ko-0.9.1')  
!./configure  
!make  
!make check  
!make install  
  
[ ] import os  
os.chdir('/tmp')  
!curl -LO https://bitbucket.org/eunjeon/mecab-ko-dic/downloads/mecab-ko-dic-2.0.1-20150920.  
!tar -zxvf mecab-ko-dic-2.0.1-20150920.tar.gz  
os.chdir('/tmp/mecab-ko-dic-2.0.1-20150920')  
!./autogen.sh  
!./configure  
!make  
# !sh -c 'echo "dicdir=/usr/local/lib/mecab/dic/mecab-ko-dic" > /usr/local/etc/mecabrc'  
!make install  
  
# install mecab-python  
import os  
os.chdir('/content')
```

```
[ ] from konlpy.tag import Mecab
```

```
[ ] # Mecab으로 토큰화 진행  
tokenizer = Mecab()
```


프로젝트 분석 절차 – 데이터 가공

“ 텍스트 벡터화 처리 ”

- 텍스트 데이터를 컴퓨터가 처리할 수 있게 수치화 변환함.
- X = 커리큘럼(전공과목)
→ 토큰화 처리
- 워드 사전

```
[ ] def make_x(x_data):  
    return tokenizer.morphs(x_data)
```

```
[ ] data.iloc[:,2].apply(make_x)
```

```
0      [건축학개론]  
1      [건축, 이론]  
2      [건축사]  
3      [건축, 설계]  
4      [건축, 구조, 의, 이해]  
  
...  
5975     [유럽, 연합, 론]  
5976     [프랑스, 외교사]  
5977     [프랑스, 문학사]  
5978     [프랑스, 시]  
5979     [프랑스, 소설]  
Name: 전공수업, Length: 5980, dtype:  
object
```

```
[ ] X_ = data.iloc[:,2].apply(make_x)
```

```
[ ] X_ # 각각 토큰화한 리스트  
  
0      [건축학개론]  
1      [건축, 이론]  
2      [건축사]  
3      [건축, 설계]  
4      [건축, 구조, 의, 이해]  
  
...  
5975     [유럽, 연합, 론]  
5976     [프랑스, 외교사]  
5977     [프랑스, 문학사]  
5978     [프랑스, 시]  
5979     [프랑스, 소설]  
Name: 전공수업, Length: 5980, dtype:  
object
```

```
[ ] X_word = [] # 모든 토큰들 (중복)  
for i in X_:  
    X_word.extend(i)
```

```
[ ] X_word #X의 단어 사전
```

```
['건축학개론',  
'건축',  
'이론',  
'건축사',  
'건축',  
'설계',  
'건축']
```

```
[ ] len(X_word)
```

```
15608
```

프로젝트 분석 절차 – 데이터 가공

“ 텍스트 벡터화 처리 ”

- 텍스트 데이터를 컴퓨터가 처리할 수 있게 수치화 변환함.
- 만든 워드 사전에 벡터 값을 부여함

```
[ ] from keras.utils.np_utils import to_categorical
```

X 카테고리

```
[ ] X_word = { v:i for i,v in enumerate(np.unique(X_word))} # X 사전
```

```
[ ] X_word
```

```
{'&': 0,  
'-': 1,  
'2': 2,  
'IT': 3,  
'UX': 4,  
'가': 5,  
'가곡': 6,  
'가공': 7,  
'가능': 8,  
'가론': 9,  
'가상': 10,  
'가상현실': 11,  
'가스': 12,  
'가요': 13,  
'가정': 14,  
'가정생활': 15,  
'가족': 16,  
'가족법': 17,  
'가창': 18,  
'각': 19,  
'각론': 20,  
...}
```

```
[ ] len(X_word)
```

2209

프로젝트 분석 절차 – 데이터 가공

“ 텍스트 벡터화 처리 “

- 텍스트 데이터를 컴퓨터가 처리할 수 있게 수치화 변환함.
- 패딩: 입력 배열의 주위를 가상의 원소로 채우는 작업. 실제 입력 값은 아니기 때문에 패딩은 “0” 으로 채운다. 계산에 영향을 미치지 않으면서 입력과 출력 데이터의 크기를 동일하게 만들 수 있다

X 패딩

```
[ ] def make_label(x):  
    x_list = []  
    for i in x:  
        x_list.append(X_word[i])  
    return x_list
```

```
X = X_.apply(make_label)
```

```
[ ] from keras.preprocessing.sequence import pad_sequences  
maxlen = X.apply(len).max()  
X = pad_sequences(X,maxlen=maxlen)
```

[] X

```
array([[ 0,  0,  0, ...,  0,  0,  60],  
       [ 0,  0,  0, ...,  0, 58, 1451],  
       [ 0,  0,  0, ...,  0,  0,  59],  
       ...,  
       [ 0,  0,  0, ...,  0, 2017, 601],  
       [ 0,  0,  0, ...,  0, 2017, 1063],  
       [ 0,  0,  0, ...,  0, 2017, 992]], dtype=int32)
```

[] X.shape

(5980, 13)

프로젝트 분석 절차 – 데이터 가공

“ 텍스트 벡터화 처리 ”

- 텍스트 데이터를 컴퓨터가
처리할 수 있게 수치화 변환함.

- y = 학과
- y 원 핫 인코딩: 텍스트를 유의미한
숫자(벡터)로 바꾸는 가장 쉬운
방법이다. N개의 단어를 각각 N차원
의 벡터로 표현하는 방식이다.
단어에 해당되는 차원(인덱스)에
1을 넣고 나머지는 0을 넣는다

```
[ ] data.학과  
0      건축학과  
1      건축학과  
2      건축학과  
3      건축학과  
4      건축학과  
...  
5975   프랑스학과  
5976   프랑스학과  
5977   프랑스학과  
5978   프랑스학과  
5979   프랑스학과  
Name: 학과, Length: 5980, dtype: object
```

```
[ ] y_ = data.학과
```

```
[ ] y_
```

```
0      건축학과  
1      건축학과  
2      건축학과  
3      건축학과  
4      건축학과  
...  
5975   프랑스학과  
5976   프랑스학과  
5977   프랑스학과  
5978   프랑스학과  
5979   프랑스학과  
Name: 학과, Length: 5980, dtype: object
```

```
[ ] y_word = { v:i for i,v in enumerate(np.unique(y_))}
```

```
[ ] y_word
```

```
{'가정관리학과': 0,  
'간호학과': 1,  
'건축설비공학과': 2,  
'건축학과': 3,  
'게임공학과': 4,  
'경영학과': 5,  
'경제학과': 6,  
'경찰행정학과': 7,  
'경호학과': 8,  
'공예학과': 9,  
'공학교육과': 10,  
'광고홍보학과': 11,  
'교육학과': 12,  
'국어국문학과': 13,  
'국제지역학과': 14,  
'국제학과': 15,  
  
'특수교육학과': 122,  
'패션디자인학과': 123,  
'프랑스어문학과': 124,  
'프랑스학과': 125,  
'한의학과': 126,  
'항공서비스과': 127,  
'항공학과': 128,  
'해양공학과': 129,  
'행정학과': 130,  
'호텔관광경영학과': 131,  
'화학공학과': 132,  
'화학과': 133,  
'환경공학과': 134}
```

```
[ ] len(y_word)
```

```
135
```


프로젝트 분석 절차 - 모델 구축 및 학습

“ 학습 데이터 구분 ”

- 만들어진 데이터를 훈련데이터와 학습데이터 구분.

```
[ ] from sklearn.model_selection import train_test_split
```

```
[ ] X_train, X_test, y_train, y_test = train_test_split(X, Y )
```

```
[ ] X_train
```

```
array([[ 0,  0,  0, ...,  0, 2065, 1661],  
       [ 0,  0,  0, ...,  0,  603, 1143],  
       [ 0,  0,  0, ...,  0,   57,   81],  
       ...,  
       [ 0,  0,  0, ...,  0, 1501, 1407],  
       [ 0,  0,  0, ...,  94,  573,  503],  
       [ 0,  0,  0, ..., 1721, 1579, 1082]], dtype=int32)
```

```
[ ] X_test
```

```
array([[ 0,  0,  0, ..., 269, 1544, 1303],  
       [ 0,  0,  0, ..., 1310, 210,  503],  
       [ 0,  0,  0, ...,   0, 1425,   24],  
       ...,  
       [ 0,  0,  0, ...,  651, 1731,  712],  
       [ 0,  0,  0, ..., 1517,  159, 1633],  
       [ 0,  0,  0, ...,  200,  165,  503]], dtype=int32)
```

```
[ ] y_train
```

```
array([[0., 0., 0., ..., 0., 0., 0.],  
       [0., 0., 0., ..., 0., 0., 0.],  
       [0., 0., 1., ..., 0., 0., 0.],  
       ...,  
       [0., 0., 0., ..., 0., 0., 0.],  
       [0., 0., 0., ..., 0., 0., 0.],  
       [0., 0., 0., ..., 0., 0., 0.]], dtype=float32)
```

```
[ ] y_test
```

```
array([[0., 0., 0., ..., 0., 0., 0.],  
       [0., 0., 0., ..., 0., 0., 0.],  
       [0., 1., 0., ..., 0., 0., 0.],  
       ...,  
       [0., 0., 0., ..., 0., 0., 0.],  
       [1., 0., 0., ..., 0., 0., 0.],  
       [0., 0., 0., ..., 0., 0., 0.]], dtype=float32)
```

```
[ ] X_train.shape, y_train.shape, X_test.shape, y_test.shape
```

```
((4485, 13), (4485, 135), (1495, 13), (1495, 135))
```

```
[ ] type(X_train)
```

```
pandas.core.series.Series
```

프로젝트 분석 절차 - 모델 구축 및 학습

“인공 신경망 구성”

- 텍스트 데이터의 연관어 분석
- 활성화 함수 : 소프트맥스

```
[ ] # 인공 신경망 구성
model = Sequential()
model.add( Embedding(2209, 8, input_length=13 ) ) # s_max =padded_x.shape[1]
model.add( Flatten() )
model.add( Dense(135, activation='softmax') )
model.summary()
```

Model: "sequential_1"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 13, 8)	17672
flatten (Flatten)	(None, 104)	0
dense (Dense)	(None, 135)	14175

=====
Total params: 31,847
Trainable params: 31,847
Non-trainable params: 0
=====

프로젝트 분석 절차 - 모델 구축 및 학습

“ 컴파일 및 학습 ”

- Early stopping callback 으로 과적합 방지

- 모델학습

- Accuracy → 62.6 %

- 모델평가

-Accuracy → 29.6 %

```
[57] 1 # 컴파일 및 학습
      2 model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
```

```
1 model.fit(X_train,
2           y_train,
3           batch_size=20,
4           epochs=200,
5           validation_data=(X_test, y_test),
6           callbacks = [early_stopping_callback] )
```

```
Epoch 32/200
225/225 [=====] - 1s 2ms/step - loss: 1.6644 - accuracy: 0.5960 - val_loss: 3.1142 - val_accuracy: 0.2936
Epoch 33/200
225/225 [=====] - 1s 3ms/step - loss: 1.6274 - accuracy: 0.6022 - val_loss: 3.1148 - val_accuracy: 0.2950
Epoch 34/200
225/225 [=====] - 1s 3ms/step - loss: 1.5924 - accuracy: 0.6076 - val_loss: 3.1172 - val_accuracy: 0.2923
Epoch 35/200
225/225 [=====] - 1s 3ms/step - loss: 1.5586 - accuracy: 0.6132 - val_loss: 3.1162 - val_accuracy: 0.2963
Epoch 36/200
225/225 [=====] - 1s 3ms/step - loss: 1.5265 - accuracy: 0.6232 - val_loss: 3.1177 - val_accuracy: 0.2957
Epoch 37/200
225/225 [=====] - 1s 3ms/step - loss: 1.4962 - accuracy: 0.6261 - val_loss: 3.1252 - val_accuracy: 0.2963
<keras.callbacks.History at 0x7f845e7672d0>
```

```
35] 1 # 예측 ->
      2 model.evaluate(X_test, y_test)
```

```
47/47 [=====] - 0s 2ms/step - loss: 3.1252 - accuracy: 0.2963
[3.1252083778381348, 0.2963210642337799]
```

프로젝트 분석 절차 – 문제점 분석과 해결 방안 도출

- 단순한 신경망 구조
- 해결: Dense 추가

```
1 # model_deep2
2 # 인공 신경망 구성
3 model_deep2 = Sequential()
4 model_deep2.add( Embedding(2209, 8, input_length=13 ) )
5 model_deep2.add( Flatten() )
6 model_deep2.add( Dense(50, activation='relu' ) )
7 model_deep2.add( Dense(50, activation='relu' ) )
8 model_deep2.add( Dense(135, activation='softmax' ) )
9 model_deep2.summary()
```

Model: "sequential_3"

Layer (type)	Output Shape	Param #
embedding_3 (Embedding)	(None, 13, 8)	17672
flatten_3 (Flatten)	(None, 104)	0
dense_5 (Dense)	(None, 50)	5250
dense_6 (Dense)	(None, 50)	2550
dense_7 (Dense)	(None, 135)	6885

=====
Total params: 32,357
Trainable params: 32,357
Non-trainable params: 0

```
Epoch 7/200
225/225 [=====] - 1s 3ms/step - loss: 4.4353 - accuracy: 0.0466 - val_loss: 5.0676 - val_accuracy: 0.0201
<keras.callbacks.History at 0x7fe912dce90>
```

```
1 # 예측 ->
2 model_deep2.evaluate(X_test, y_test)
```

```
47/47 [=====] - 0s 2ms/step - loss: 5.0676 - accuracy: 0.0201
[5.067601203918457, 0.020066889002919197]
```


프로젝트 분석 절차 – 문제점 분석과 해결 방안 도출

- 모델 변경
- 해결: LSTM

```
1 model2 = Sequential()  
2 model2.add( Embedding(2209, 8, input_length=13 ) )  
3 model2.add(LSTM(13, activation='tanh')) # LSTM tanh사용 -1부터 1사이 시그모이드는 0~2  
4 model2.add(Dense(135, activation='softmax'))  
5 model2.summary()
```

```
225/225 [=====] - 2s 10ms/step - loss: 4.6664 - accuracy: 0.0321 - val_loss: 4.8230 - val_accuracy: 0.0161  
<keras.callbacks.History at 0x7fe915871350>
```

✓
3초

```
1 # 예측 ->  
2 model2.evaluate(X_test, y_test)
```

```
47/47 [=====] - 17s 373ms/step - loss: 4.8230 - accuracy: 0.0161  
[4.822968006134033, 0.016053510829806328]
```

프로젝트 분석 절차 – 문제점 분석과 해결 방안 도출

1 z_word

- 부족한 데이터
- 적은 input 많은 output
- 해결: 데이터 추가
- 해결: 예측 output 줄이기
- 분류가 적은 '계열' 예측

{'공학계열': 0, '교육계열': 1, '사회계열': 2, '예체능계열': 3, '의약계열': 4, '인문계열': 5, '자연계열': 6}

	계열	학과	전공수업
0	공학계열	건축학과	건축학개론
1	공학계열	건축학과	건축이론
2	공학계열	건축학과	건축사
3	공학계열	건축학과	건축설계
4	공학계열	건축학과	건축구조의 이해
...
5975	인문계열	프랑스학과	유럽연합론
5976	인문계열	프랑스학과	프랑스 외교사
5977	인문계열	프랑스학과	프랑스 문학사
5978	인문계열	프랑스학과	프랑스 시
5979	인문계열	프랑스학과	프랑스 소설

5980 rows × 3 columns

프로젝트 분석 절차 – 문제점 분석과 해결 방안 도출

```
[140] 1 # 인공 신경망 구성
      2 model = Sequential()
      3 model.add( Embedding(2209, 8, input_length=13 ) )
      4 model.add( Flatten() )
      5 model.add( Dense(7, activation='softmax') )
```

```
[143] 1 model.summary()
```

Model: "sequential_16"

Layer (type)	Output Shape	Param #
embedding_16 (Embedding)	(None, 13, 8)	17672
flatten_14 (Flatten)	(None, 104)	0
dense_34 (Dense)	(None, 7)	735
Total params: 18,407		
Trainable params: 18,407		
Non-trainable params: 0		

```
Epoch 17/200
225/225 [=====] - 1s 3ms/step - loss: 0.3535 - accuracy: 0.8979 - val_loss: 0.9147 - val_accuracy: 0.7164
Epoch 18/200
225/225 [=====] - 1s 3ms/step - loss: 0.3366 - accuracy: 0.8990 - val_loss: 0.9216 - val_accuracy: 0.7137
Epoch 19/200
225/225 [=====] - 1s 3ms/step - loss: 0.3214 - accuracy: 0.9032 - val_loss: 0.9273 - val_accuracy: 0.7117
<keras.callbacks.History at 0x7fe91587a990>
```

```
1 model.evaluate(X_test, Z_test)
```

```
47/47 [=====] - 0s 2ms/step - loss: 0.9273 - accuracy: 0.7117
[0.9272875785827637, 0.7117056846618652]
```

학습정확도: Accuracy92%

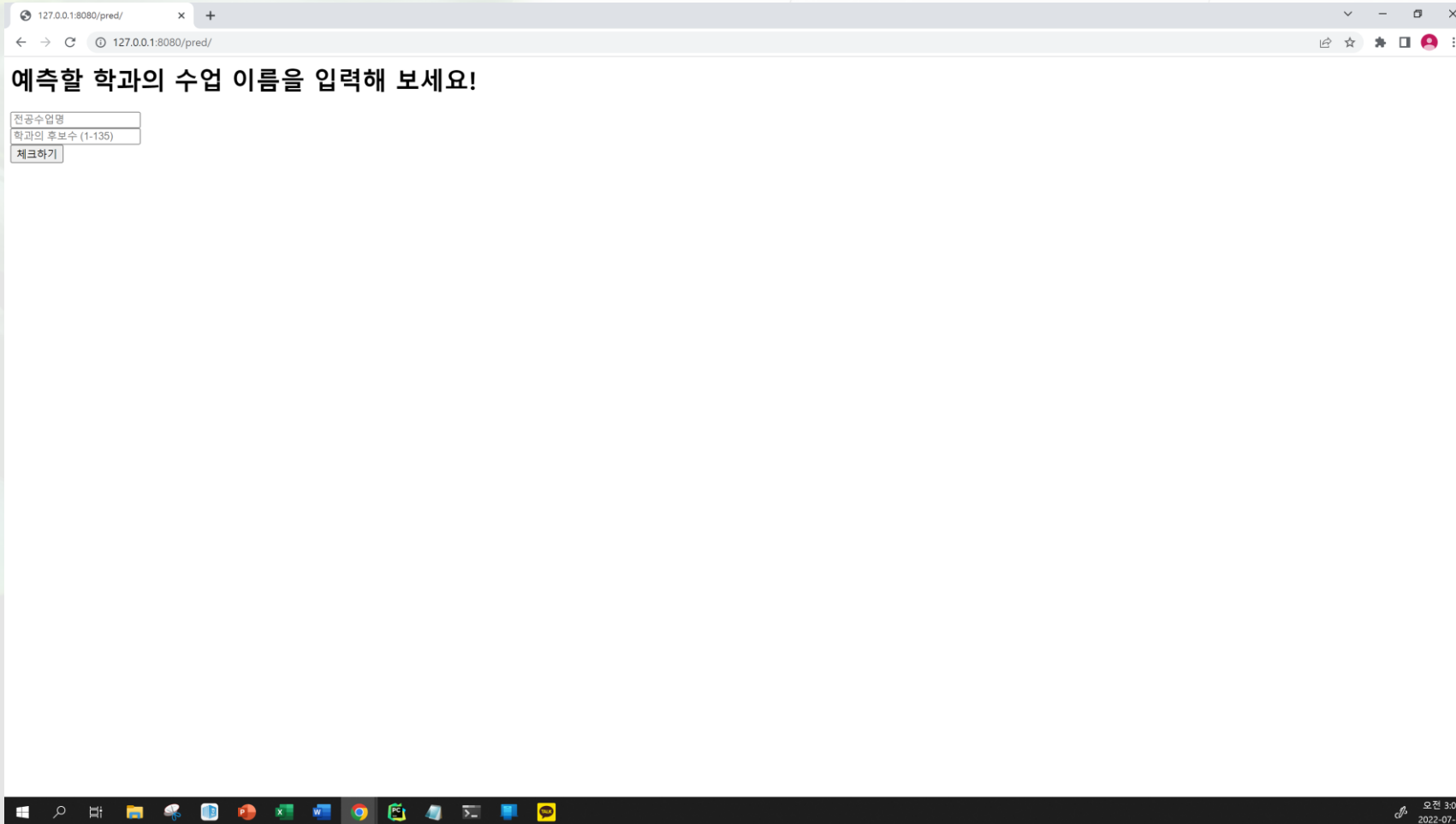
평가정확도 :Accuracy71%

프로젝트 분석 절차 – 문제점 분석과 해결 방안 도출

- 형태소 분석기
- 단어 길이가 너무 짧음
- 사전에 없는 복합어 형태가 많음
 - 정확한 의미단위의 형태소 분석 어려움
- 해결: 다른 형태소 분석기 사용
- 해결: 데이터 띄어쓰기 추가

```
['종교사']  
['불교', '개론']  
['불교', '윤리학']  
['한국', '불', '교사']  
['중국', '불', '교사']  
['인도', '불', '교사']
```


프로젝트 구현 웹 -django



The screenshot shows a web browser window with the address bar displaying "127.0.0.1:8080/pred/". The page content includes the heading "예측할 학과의 수업 이름을 입력해 보세요!" (Please enter the name of the course you want to predict!). Below the heading, there are two input fields: "전공수업명" (Major Course Name) and "학과의 후보수 (1-135)" (Candidate number of the department (1-135)). A "체크하기" (Check) button is located below the second input field. The browser's taskbar at the bottom shows various application icons and the system clock indicating "오전 3:09 2022-07-15".

127.0.0.1:8080/pred/

← → ↻ 127.0.0.1:8080/pred/

예측할 학과의 수업 이름을 입력해 보세요!

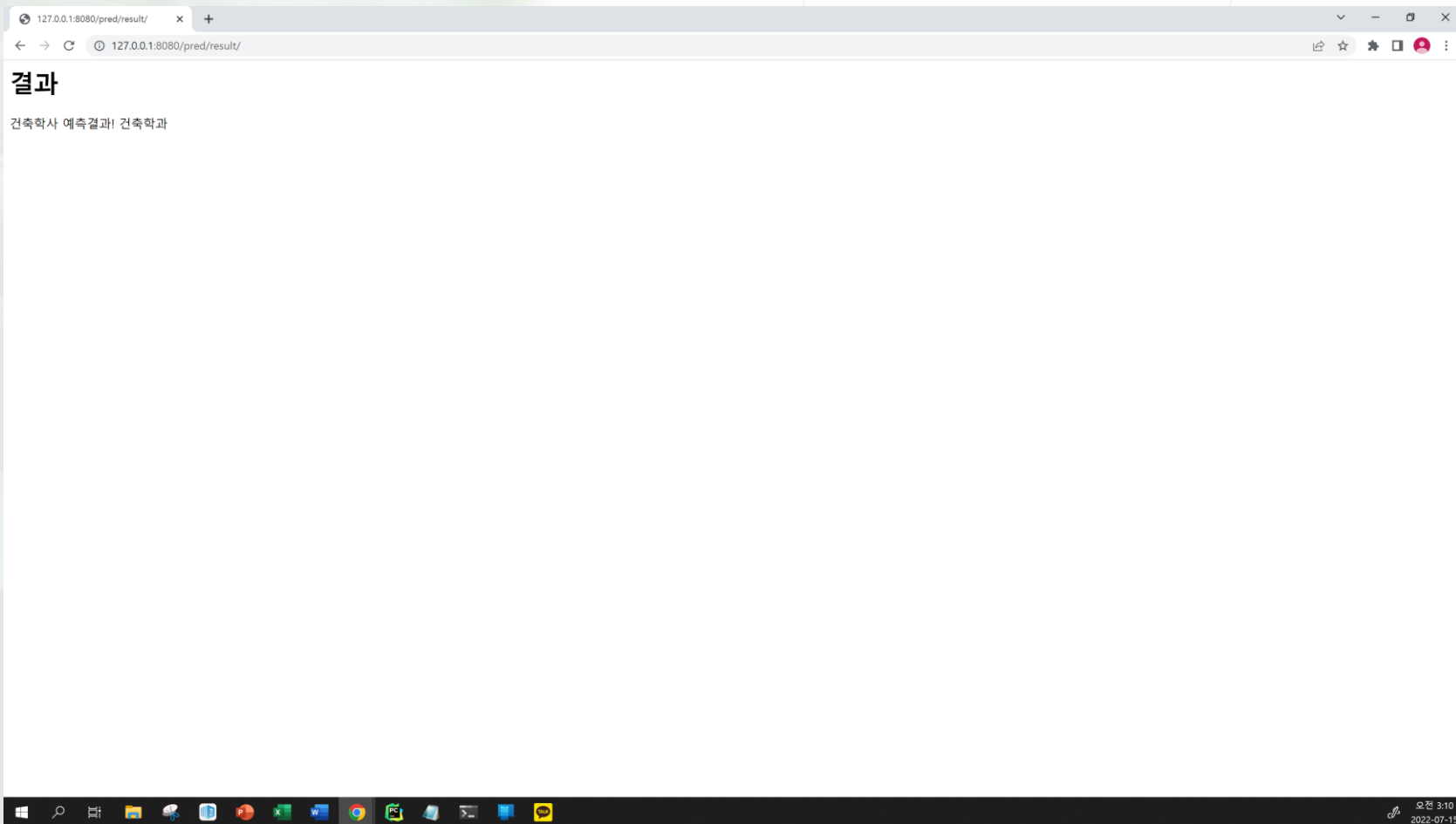
전공수업명

학과의 후보수 (1-135)

체크하기

오전 3:09 2022-07-15

프로젝트 구현 웹 -django



결과 도출 및 평가

이번 연구는 대학교 전공 교과목 커리큘럼을 토대로, 관련 교과목이 어떤 학과들과 가장 연관성이 있는지에 대한 연구를 진행하였다.

전공 교과목의 입력값(변수)를 더 많이 확보 한다면, Accuracy 값을 좀 더 높일 수 있으며 최적화된 학과 추천 모델을 구현할 수 있을 것이다.



맺음말...



데이터 확장성

대학교에 개설된 학과별 커리큘럼을 토대로 관련 교과목이 어떤 학과들과 가장 연관성이 있는지에 대한 연구를 진행하였다.

이러한 데이터를 기반으로 졸업 후 진로 데이터와 접목 시킨다면 좀 더 유의미한 학과 추천 모델을 제안할 수 있을 것이다.



시간적 제약

해당 프로젝트를 진행하면서 제한된 시간적 제약으로 보다 완성도 있는 프로젝트 결과물을 만들지 못한 부분에 대해 많은 아쉬움이 남는다.

하지만 프로젝트를 진행하면서 텍스트 마이닝에 대한 개념과 무한한 효용성에 대해서 알 수 있었던 좋은 기회였다.

참고문헌/ 자료

* 인용 논문, 자료

1. 한국교육개발원 2021 학과(전공) 분류 자료집
이기준 & 오지연 & 이정환, 김은미, 구경아, 설가인, 이윤정, 오연서, 이경록
2. 교과 연계 진로 탐색을 위한 인공지능 기반 고교 선택교과 및 대학 학과 추천 시스템
백진현 & 김하연 & 권기원 (2020.11)
3. 한국어의 내용분석을 위한 KrKwic 프로그램의 이해와 적용
박한우 & Loet Leydesdorff (2004,10)

* 인용 웹사이트 :

<https://www.work.go.kr/consltJobCarpa/srch/schdpt/schdptSrch.do>

<http://www.slideshare.net/hanpark>

<http://www.slideshare.net/hanpark/krkwicnovember2006>

감사합니다

대학 교과목 분석을 통한
학과 추천 모델 연구

- 1 조 -