



선배님 수강과목을 부탁드립니다

3조 이동호 김아영 박재영 임세정

Contents

01

프로젝트 소개

- 1.1. 수요기업 과제개요
- 1.2. 팀원소개
- 1.3. WBS
- 1.4. 프로젝트 추진배경

02

데이터 분석

- 2.1 데이터 소개
- 2.2. 데이터 전처리
- 2.3. 전처리 완료 데이터

03

수강 과목 추천 과정 및 결과

- 3.1. 분석방법
- 3.2. Model1 입력변수 & 목표변수 설명
- 3.3. Model1 분석결과
- 3.4. Model2 분석결과 (서브모델)

04

서비스 도출 결과

- 4.1. 서비스 도출 결과
- 4.2. 활용방안 및 기대효과
- 4.3. 한계점 및 개선점



프로젝트 소개

- 1.1. 수요기업 과제개요
- 1.2. 팀원소개
- 1.3. WBS
- 1.4. 프로젝트 추진배경

1.1. 수요기업 과제 개요

과제목표

성적정보와 이력서정보를
기반으로 한 **수강과목 추천 서비스**

프로젝트 요건정의

대상

대학교에서 수강신청 가이드라인이 필요한 학생
특정 직무에 도움이 되는 수강과목을 추천받고 싶은 대학생

프로젝트 배경

팬데믹으로 야기된 비대면 시대에 끊어진 인적네트워크를 대신할 수 있는
연결매개체가 필요
대나무숲 ~ 에브리타임으로 이어진 대학 커뮤니티 플랫폼의 영향력 증가

프로젝트 요구사항

학생 개인들의 성적, 희망 직무를 활용하여 수강과목을 추천
추후 발전 가능한 프로토타입모델 개발

기대효과

기대효과

인적네트워크의 의존도 높았던 수강과목
선정을, AI가 대신 해주어 효율성을 높임

직무관련도가 높은 수강과목 추천으로 취
업에 도움

기존의 증명서 발급 서비스에 더해 수강과
목 큐레이션 서비스까지 사업 영역을 확장,
맞춤형 인재 양성을 통한 부가 가치를 창출

1.2. 프로젝트 팀원 소개

선배님
수강과목을 부탁드립니다



이동호 기획

- 프로젝트 기획
- 방향설정
- 스케줄 조정 및 발표



김아영 분석/개발

- 데이터셋 가공
- 데이터 전처리
- 머신러닝 모델 개발



임세정 분석/개발

- 데이터 EDA
- 데이터 전처리
- 머신러닝 모델 개발



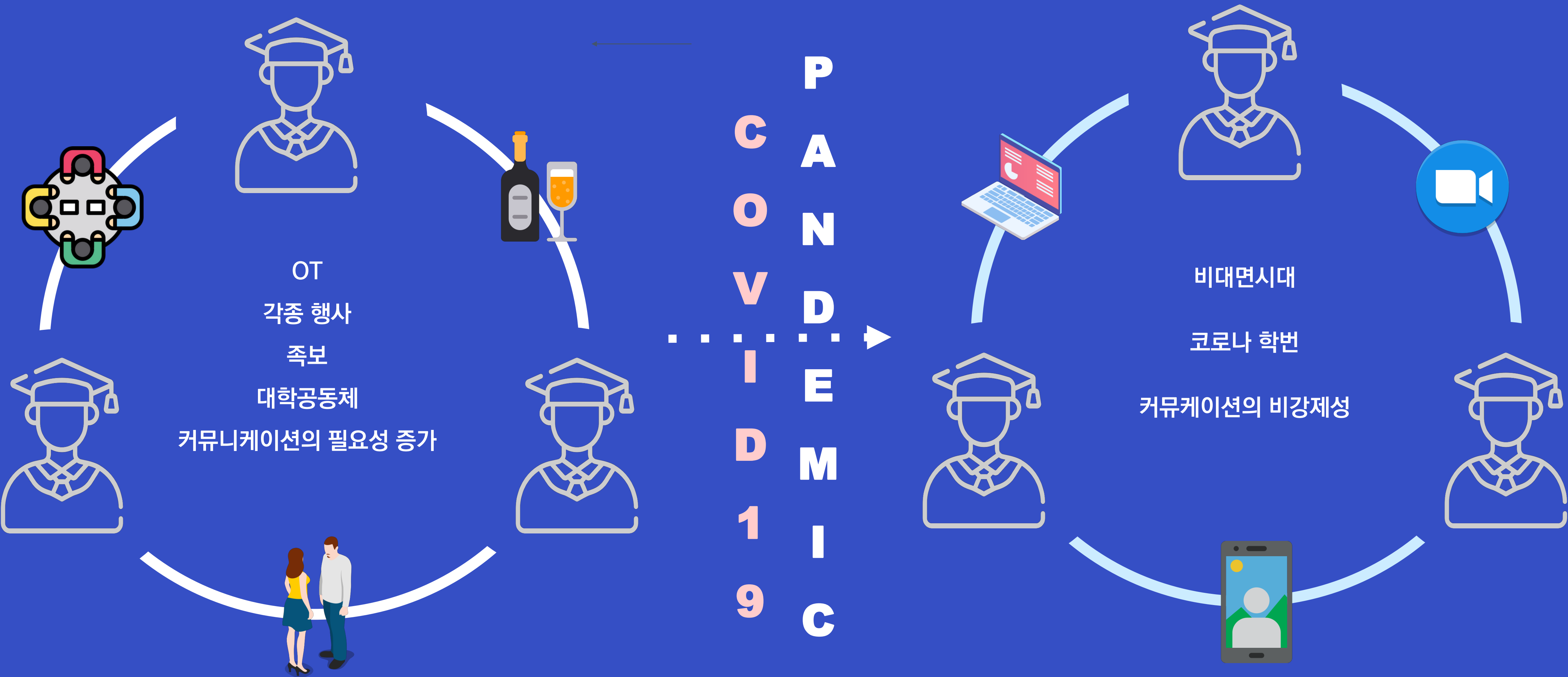
박재영 분석/개발

- 모델 개발
- 모델 고도화
- 모델 테스트

1.3. WBS (일정 계획)

[illegible]

1.4. 프로젝트 추진배경



1.4. 프로젝트 추진배경

친구도 스펙도 못 만들고 2년째 '줌 수업'... 잃어버린 낭만 [심
충기획]

“아직도 캠퍼스가 낯설어요” 코로나19 팬데믹 속 혼란 겪는 20·21학번 대학생들

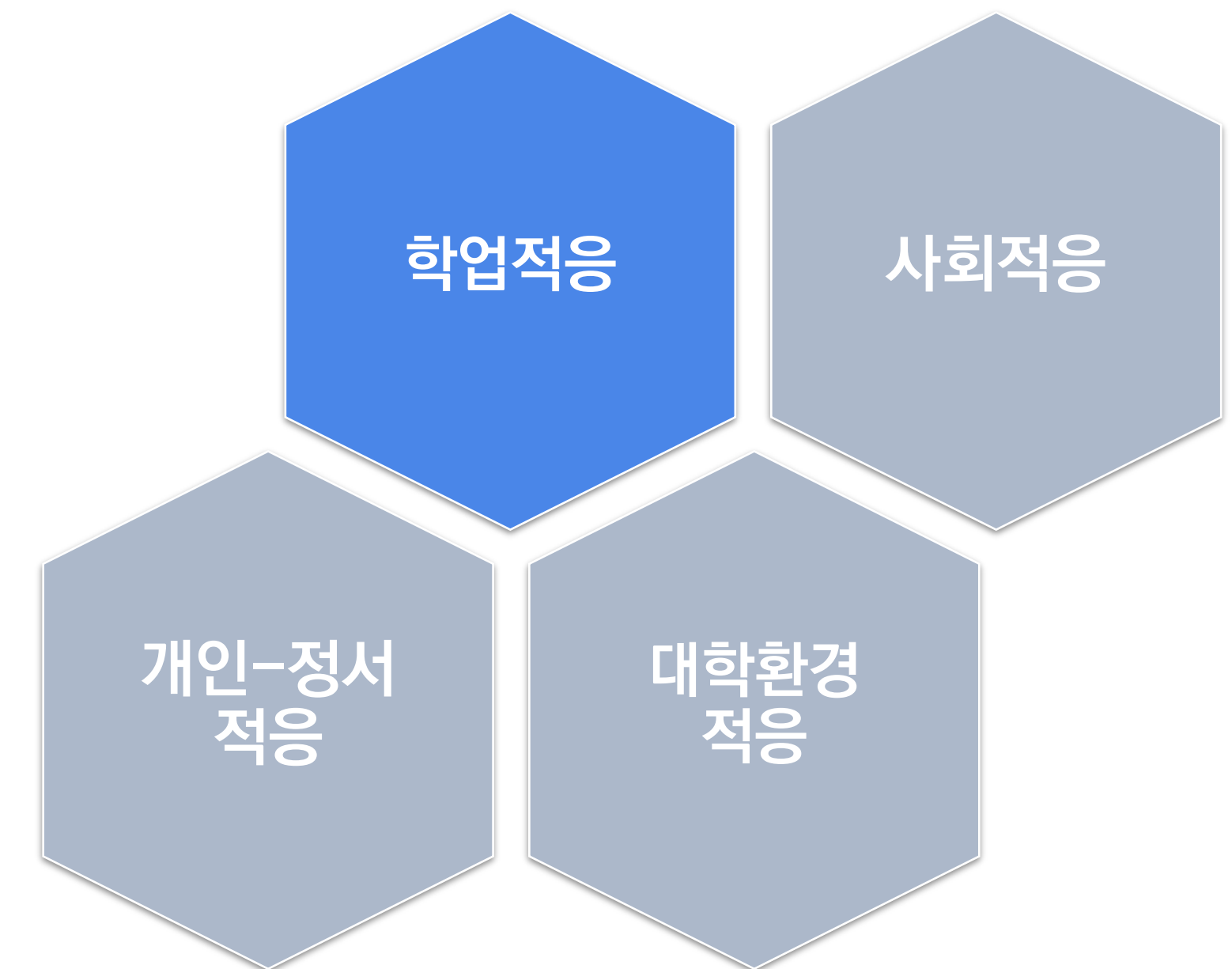
코로나 학번의 빼앗긴 봄

카페서 대학생할 시작... “과방이 뭐예요” 코로나학번의 눈물

정부, 코로나19로 위기 놓인 청년층 위해 특별대책 수립...취업·주거·등록금 지원



베이커와 시리크의
대학생활적응척도(SACQ)



참고) 이종만. (2020). 코로나19 원격 교육에서 외로움과 유튜브
과다사용이 대학생활적응에 미치는 영향에 대한 탐색적 연구



데이터 분석

- 2.1. 데이터 소개
- 2.2. 데이터 전처리
- 2.3. 전처리 완료 데이터

2.1. 데이터 소개

| 순번 | 성별 | 학교명 | 단과명 | 주전공 | 복수전공 | 부전공 | 입학일 | 졸업일 |
|-------|----|---------|----------|--------|-------|-----|----------|----------|
| A0001 | 여 | OO대학교 | | 사회체육학과 | | | 19930311 | 20010227 |
| A0002 | 남 | OO대학교 | 자연과학대학 | 수학과 | | | 35187 | 38571 |
| A0003 | | OO여자대학교 | 공과대학 | 자원공학과 | X | X | 19940302 | 19970221 |
| A0004 | 여 | OO대학교 | | 정보통계학과 | | | 33039 | 37162 |
| A0005 | 여 | OO여자대학교 | 전자정보공과대학 | 컴퓨터공학 | X | X | 37319 | 40939 |
| A0006 | 남 | OO대학교 | | 디자인과 | X | X | 19980301 | 20020219 |
| A0007 | | OO대학교 | | 컴퓨터공학과 | | | 39820 | 43029 |
| A0008 | 남 | OO대학교 | | 아동학과 | | | 20020303 | 20080203 |
| A0009 | 남 | OO대학교 | | 경영학 | 회계학전공 | X | 39210 | 45938 |
| A0010 | 여 | OO대학교 | | 전자과 | | | | |



| 순번 | 년도 | 학기 | 이수구분 | 교과목명 | 성적 | 학점 |
|-------|------|----|------|---------|----|----|
| A0001 | 1996 | 1 | 교양 | 채플 | P | P |
| A0002 | 1998 | 2 | 전공 | 운동보건학 | B | 2 |
| A0003 | 1997 | 2 | 전공 | 스포츠사회학 | A- | 3 |
| A0004 | 1998 | 1 | 교양 | 세계정치의이해 | A0 | 3 |
| A0005 | 1998 | 1 | 전선 | 씨름 | B+ | 1 |
| A0006 | 1995 | 1 | 교선6 | 레크리에이션 | A | 2 |
| A0007 | 1994 | 2 | 전선 | 사회체육강독 | A+ | 3 |
| A0008 | 1992 | 2 | 전공 | 스포츠의학 | C | 3 |
| A0009 | 1995 | 2 | 전공 | 선형대수 | B | 2 |



| 순번 | 근무경력 |
|-------|--|
| A0001 | -2004년 11월~ 2005월 현재: OO공단 OO지도원 근무 / 일반사무 담당 - 사업계획서 작성 |
| A0002 | |
| A0003 | |
| A0004 | 2002년2월17일~2004년4월 6일 (주)OO기업 e-biz사업부 개발팀 근무 |
| A0005 | |
| A0006 | |
| A0007 | |
| A0008 | 거래처 관리 및 신규 거래처 개척 |
| A0009 | 디자인 담당업무를 진행했습니다. 의류 샘플 선정부터 디자인에 관해 모든 업무 처리 |
| A0010 | OO기업 마케팅팀 8개월 근무 경력은 많지 않지만 성실하게 근무 가능합니다! |



학생
1675명

(실제 데이터 기반으로 구성된 가상의 데이터입니다)

2.2. 데이터 전처리

순번

- 중복되는 순번 제거
- A0001 ~ A2162 (1675명)

성적

- 학교별로 표기의 차이
- A+부터F까지 9~0 으로 분류
- 오류값 (예: AB+, /, S, AD) 처리
- (raw데이터에서 범주개수 : 87 -> 10개)

단과명

- 학교별로 다양한 단과명이 존재
- 유사한 단과명의 경우, 하나로 분류
- 결측치는 주전공을 보고 분류
- (raw데이터에서 범주개수 :139 -> 15개)

학점

- 수업의 방식에 따라 학점의 차이
- 실습 등을 제외한 1, 2, 3 학점만 사용
- (raw데이터에서 범주개수 : 60 -> 3개)

이수구분

- 학교별로 표기의 차이
- '전필', '전선', '교양' 3가지로 구분
- 표기오류의 경우 직접 과목을 보며 분류
- (raw데이터에서 범주개수 : 331 -> 3개)

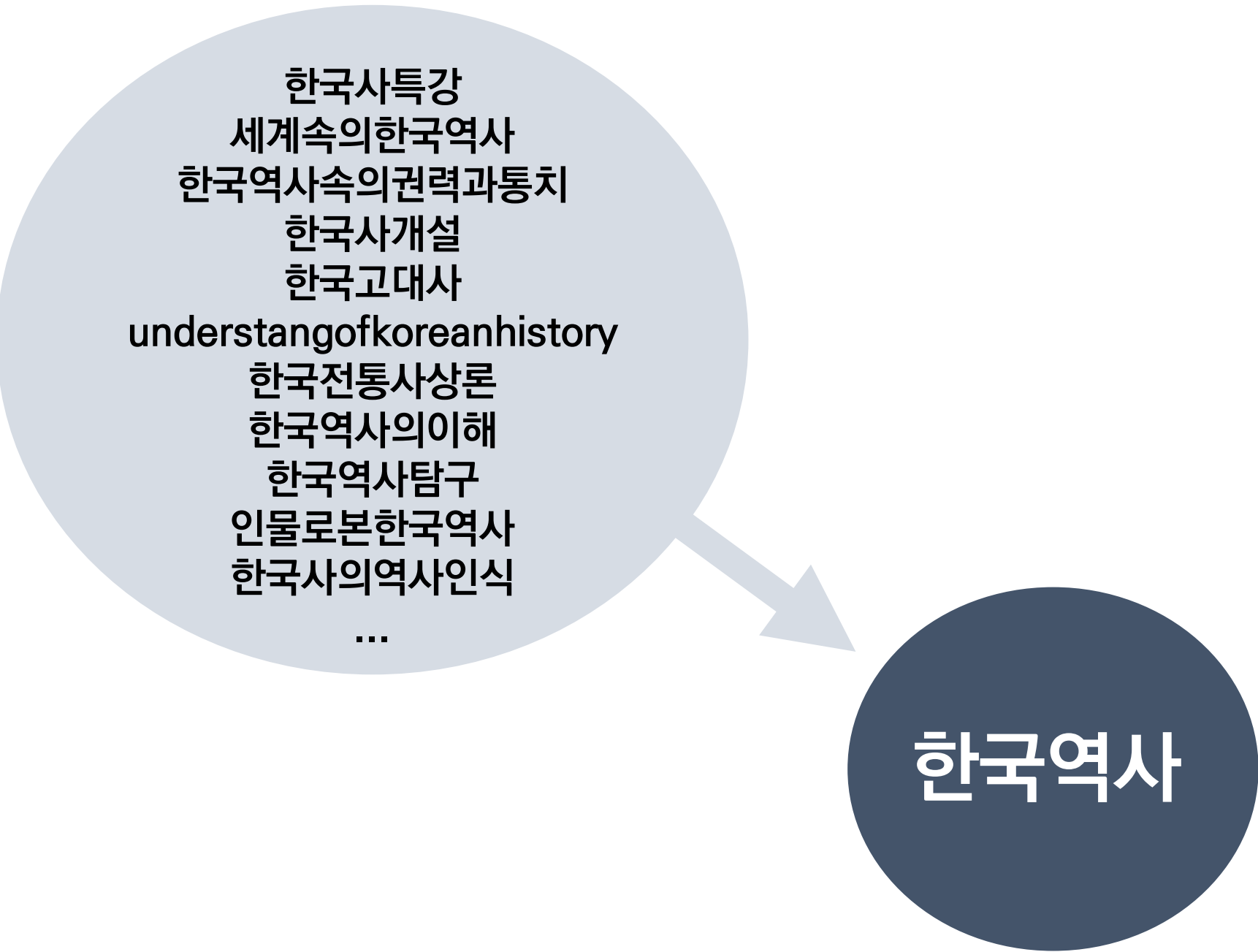
전체학기

- 년도와 학기컬럼을 이용해 학기를
순서에 따라 1~8학기로 구분
- 9학기 이상은 데이터에서 제외

2.2. 데이터 전처리

교과목명

- 2만개 이상의 교과목명을 100개로 라벨링 작업
- raw데이터에서 범주개수 : 23953



경력사항

- 경력데이터와 전공 확인 후 '직무' 도출작업
- NCS기반 중분류까지 라벨링 (약 40가지)

| 순번 | 경력사항 |
|-------|-------------------------------|
| A0011 | OO기업 의류디자이너 스포츠 아웃도어 의류디자이너 |

| 코드 | NCS 직무 상세 |
|-----|-------------------------|
| 01 | 관리직(임원·부서장) |
| 02 | 경영·행정·사무직 |
| 024 | 광고·조사·상품기획·행사기획 전문가 |
| 026 | 경영지원 사무원 |
| 028 | .무역·운송·생산·품질 사무원 |
| 029 | 안내·고객상담·통계·비서·사무보조 및 기타 |
| 03 | 금융·보험직 |
| | ... |
| 41 | 예술·디자인·방송직 |
| 42 | 스포츠·레크리에이션직 |
| | ... |
| 88 | 인쇄·목재·공예 및 기타 설치·정비·생산직 |
| 89 | 제조 단순직 |
| 90 | 농림어업직 |

2.3. 최종 데이터셋

초기 데이터셋

| | 순번 | 이수구분 | 교과목명 | 성적 | 학점 | semester | 단과명 | 직무중분류 |
|-------|-------|------|-------|-----|-----|----------|--------|-------|
| 0 | A0001 | 교양 | 철학 | 9 | 2 | 6 | 체육대학 | 26.0 |
| 1 | A0001 | 전선 | 스포츠일반 | 7 | 3 | 5 | 체육대학 | 26.0 |
| 2 | A0001 | 전선 | 스포츠일반 | 8 | 3 | 5 | 체육대학 | 26.0 |
| 3 | A0001 | 전선 | 스포츠기타 | 9 | 2 | 5 | 체육대학 | 26.0 |
| 4 | A0001 | 전선 | 스포츠일반 | 7 | 2 | 5 | 체육대학 | 26.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 50656 | A2162 | 교양 | 동양학 | 9 | 3 | 4 | 생활과학대학 | 24.0 |
| 50657 | A2162 | 교양 | 예술학 | 8 | 3 | 4 | 생활과학대학 | 24.0 |
| 50658 | A2162 | 전선 | 패션기타 | 8 | 2 | 4 | 생활과학대학 | 24.0 |
| 50659 | A2162 | 교양 | 영어회화 | 7 | 2 | 5 | 생활과학대학 | 24.0 |
| 50660 | A2162 | 전선 | 패션기타 | 6 | 3 | 2 | 생활과학대학 | 24.0 |

A0001

A2162

최종 학습용 데이터셋

| | 순번 | 단과명 | 직무중 분류 | 1_1교과목 명 | 1_1성 적 | 1_1이수 구분 | 1_1학 점 | 1_2교과 목명 | 1_2성 적 | 1_2이수 구분 | ... |
|------|-------|------------|-----------|--------------|-----------|-------------|-----------|-------------|-----------|-------------|-----|
| 0 | A0001 | 체육대학 | 26 | 스포츠일반 | 7 | 전선 | 3 | 구기스포 츠 | 7 | 전선 | ... |
| 1 | A0004 | 공과대학 | 13 | 산업일반 | 7 | 교양 | 2 | 생활학 | 8 | 교양 | ... |
| 2 | A0005 | 공과대학 | 61 | 물리학 | 7 | 교양 | 2 | 화학 | 7 | 교양 | ... |
| 3 | A0006 | 공과대학 | 13 | 구기스포츠 | 6 | 교양 | 1 | 철학 | 7 | 전선 | ... |
| 4 | A0009 | 공과대학 | 61 | 컴퓨터소프 트웨어 | 5 | 전선 | 1 | 영어회화 | 5 | 교양 | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1059 | A2157 | 상경대학 | 26 | 경영학 | 8 | 전선 | 3 | 커뮤니케 이션 | 8 | 교양 | ... |
| 1060 | A2159 | 자연과학 대학 | 29 | 영어회화 | 4 | 교양 | 3 | 국문학 | 6 | 교양 | ... |
| 1061 | A2160 | 사회과학 대학 | 24 | 회계학 | 6 | 전선 | 2 | 통계학 | 8 | 전선 | ... |
| 1062 | A2161 | 사회과학 대학 | 2 | 한국역사학 | 2 | 교양 | 3 | 경제학 | 5 | 전선 | ... |
| 1063 | A2162 | 생활과학 대학 | 24 | 국문학 | 6 | 교양 | 3 | 환경학 | 8 | 교양 | ... |

3

수강과목 추천 과정 및 결과

3.1. 분석방법

3.2. Model1 입력변수 & 목표변수 설명

3.3. Model1 분석결과

3.4. Model2 분석결과 (서브모델)

3.1. 분석 방법

| 입력변수 설정 | 목표변수 설정 | ML 모델설정 | 파라미터설정 | 모델 테스트 |
|---|-------------------------------------|--|---|---|
| <div>학생의 희망직무</div> <div>전공 (단과명)</div> <div>1~2학년 수강과목</div> <div>이수구분</div> <div>성적</div> <div>학점</div> | Model1(메인) | <div>[분류모델]</div> <div>Decision Tree</div> <div>Catboost</div> <div>RandomForest</div> <div>LightGBM</div> | Pipeline | <div>Accuracy</div> <div>F1score</div> <div>ROC AUC</div> |
| | 3학년 1,2학기 / 4학년 1,2학기 추천 수강과목 | | MInMaxScaler Gridsearch Label Encoding SMOTE(Oversampling) | |
| | Model2(서브) | | Depth Learning rate Iterations | |
| 3~4학년 과목별 수강여부 | | | | |

3.2. Model1 입력변수 & 목표변수 설정



3.3. Model1 분석결과

모델 선정
배경

- 상대적으로 데이터가 많은 3학년 1학기의 경우 빠르고 성능이 좋은 Light GBM 선정
- 학기가 높아질수록 듣는 수강과목이 줄어들어 Row의 수가 감소. 과적합을 방지하기 위하여 Random Forest 선정

성능 향상
작업

- Label Encoding
- Pipeline
- MinMax Scalar
- SMOTE 기법 사용
- GridSearch, Cross Validation

| 〈추천대상학기〉 | Row 개수 | Best Model | Train Data | | Test Data | |
|----------|--------|---------------|------------|----------|-----------|----------|
| | | | ACC | F1 score | ACC | F1 score |
| 3학년1학기 | 277개 | Light GBM | 79% | 78% | 46% | 42% |
| 3학년2학기 | 256개 | Random Forest | 89% | 90% | 54% | 47% |
| 4학년1학기 | 222개 | Random Forest | 93% | 93% | 43% | 48% |
| 4학년2학기 | 166개 | Random Forest | 98% | 98% | 47% | 45% |

3.4. Model2 분석결과 (서브모델)

01

목표변수 설정

- 각 교과목을 목표변수로 설정 > 과목별로 모델 설정 및 학습
- 목표변수 : 3~4학년에 해당 과목을 수강한 경우 1 , 아닌 경우 0
- 교과목선정 기준 : 200명 이상이 수강한 과목을 선정

02

Over sampling

- SMOTE 기법 활용

03

Catboost 학습 모델 사용

- 이진분류 모델에서 가장 우수한 성능을 내어 채택

04

학습모델 정확도 평가

- f1score
- ROC - AUC

| 교과목명 | 3학년 이후 수강한 학생(명) | 임계값 | Train data | Test data | |
|----------|---------------------|-----|--------------|--------------|-----------|
| | | | f1 macro avg | f1 macro avg | ROC - AUC |
| 경영학 | 365 | 0.5 | 0.92 | 0.60 | 0.6375 |
| 경제학 | 302 | 0.5 | 0.90 | 0.64 | 0.6929 |
| 미술학 | 294 | 0.5 | 0.84 | 0.58 | 0.6336 |
| 컴퓨터일반 | 269 | 0.5 | 0.92 | 0.72 | 0.7840 |
| 국제사 | 265 | 0.4 | 0.87 | 0.52 | 0.5158 |
| 스포츠일반 | 256 | 0.5 | 0.88 | 0.48 | 0.5308 |
| 국문학 | 253 | 0.5 | 0.83 | 0.57 | 0.6446 |
| 영어회화 | 243 | 0.5 | 0.94 | 0.56 | 0.6038 |
| 미디어 | 239 | 0.5 | 0.87 | 0.57 | 0.5963 |
| 가정학 | 238 | 0.4 | 0.89 | 0.50 | 0.5107 |
| 교육학 | 226 | 0.5 | 0.83 | 0.46 | 0.5305 |
| 사회학 | 226 | 0.5 | 0.85 | 0.57 | 0.6637 |
| 비즈니스 | 215 | 0.5 | 0.87 | 0.68 | 0.7165 |
| 철학 | 214 | 0.5 | 0.93 | 0.54 | 0.5942 |
| 컴퓨터소프트웨어 | 208 | 0.3 | 0.95 | 0.59 | 0.6728 |



서비스 도출 결과

- 4.1. 서비스 도출 결과
- 4.2. 활용방안 및 기대효과
- 4.3. 한계점 및 개선점

4.1. 서비스 도출 결과 - 프로토타입



OO 대학교 수강과목 추천 서비스

다음 사항을 입력하세요 :

학년, 학기: 4학년 1학기

단과명: 자연과학대학

희망 직무: 섬유·의복 생산직

- ...
- 보건/의료직
- 예술·디자인·방송직

성적표 첨부: ☐ 첨부

4-1학기 추천과목

- 1순위: 교육학
- 2순위: 경제학
- 3순위: 컴퓨터일반



과목 정보

| | | | |
|--------------|-----------|-----|------------------|
| 현대교육 사상 | 임OO 교수 | 월수금 | 9:00 ~ 9:50 |
| 인간학습 과 발달 | 박OO 교수 | 화목 | 11:00 ~ 12:20 |
| ... | ... | ... | ... |

```
Y.unique()
array(['경영학', '경제학', '교육학', '국문학', '미술학', '컴퓨터일반', '패션기타', '패션일반'],
      dtype=object)

Y.unique()
array([ 4,  5, 11, 13, 29, 84, 88, 91])

best_model.predict_proba(check)
array([[0.07428571, 0.18428571, 0.42761905, 0.06, 0.04,
        0.08857143, 0.0452381, 0.08]])
```

4.1. 서비스 도출 결과 - 최종형태 예시



이름

라이언

학번

201925127

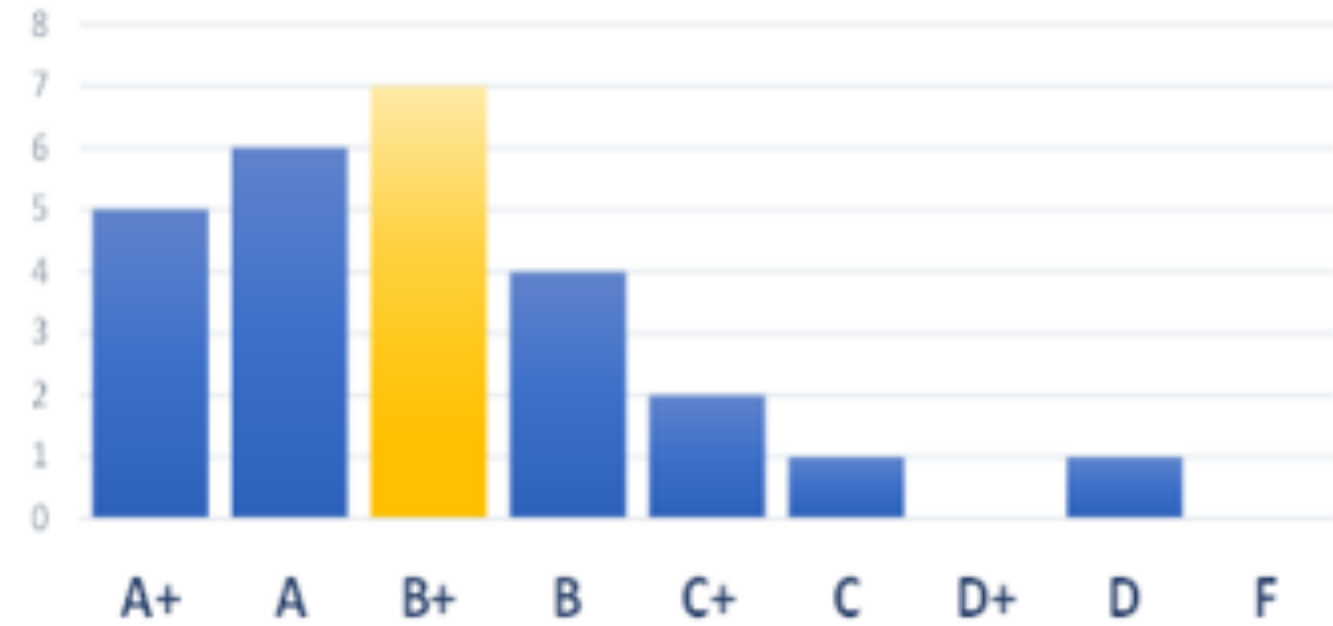
학과

통계학과

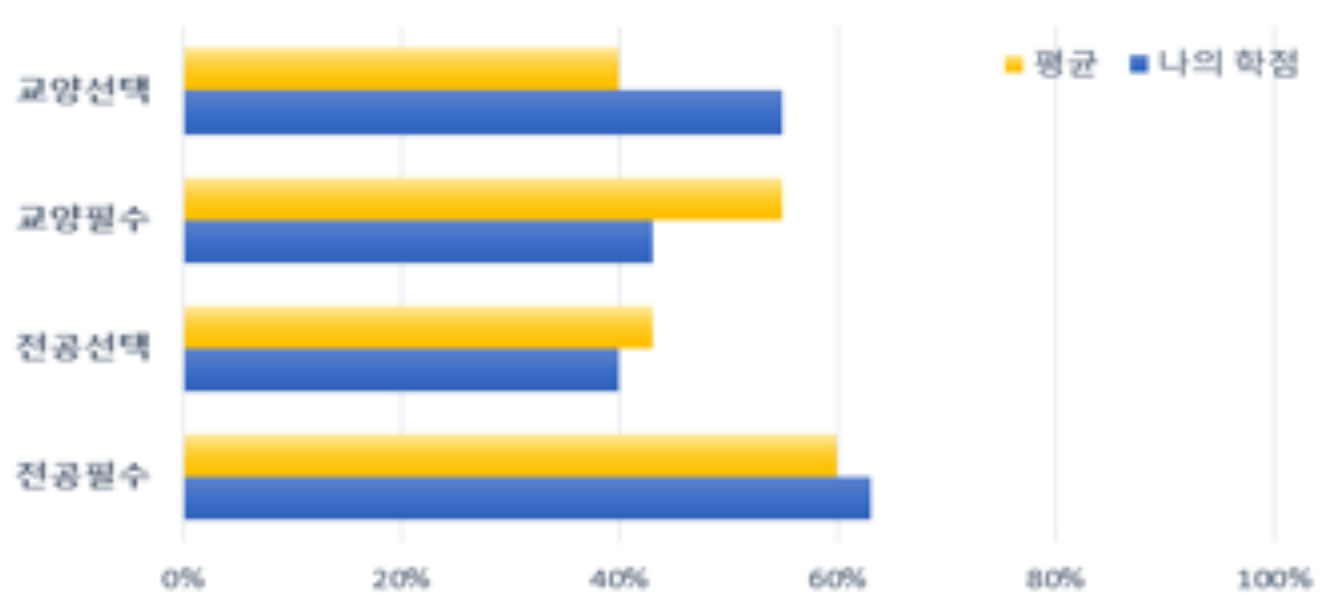
희망직무

데이터분석가

나의 성적분포



분야별 이수한 학점



AI가 추천하는 다음학기 수강과목

| | 1학년 수강과목 | 2학년 수강과목 | 3학년 1학기 추천 수강과목 | 3학년 2학기 |
|----|---|--|--|--|
| 전공 | <div>1-1 응용 수학1 기초통계학1 통계학원론 미분과 적분1</div> <div>1-2 응용 수학2 기초통계학2 정보통계학 선형대수학</div> | <div>2-1 기초확률론 보험학원론 탐색적 자료분석</div> <div>2-2 행렬대수학 미시경제학 재무관리 전산실습</div> | <div>전공 필수 회귀분석 통계적 품질혁신론</div> <div>전공선택 다변량자료 분석 시계열분석</div> | <div>전공 필수 비모수통계학</div> <div>전공선택 통계적 인공지능</div> |
| 교양 | <div>1-1 실용컴퓨터1 CSP 진로탐색 실용영어1</div> <div>1-2 실용컴퓨터2 경영과 경제</div> | <div>2-1 현재중국론 프랑스 예술과 광기</div> <div>2-2 아메리칸재즈 테니스 기초</div> | <div>교양필수 경영학원론 정치학개론</div> <div>교양선택 빅데이터 마케팅 비즈니스 커뮤니케이션</div> | <div>교양필수 금융통계학</div> <div>교양선택 AI와 미래산업</div> |

수강한 과목의 희망직무적합도

good! 73%

희망직무 채용공고

CACAO BANK
[프로덕트] 데이터 분석가 모집

4.2. 활용방안 및 기대효과

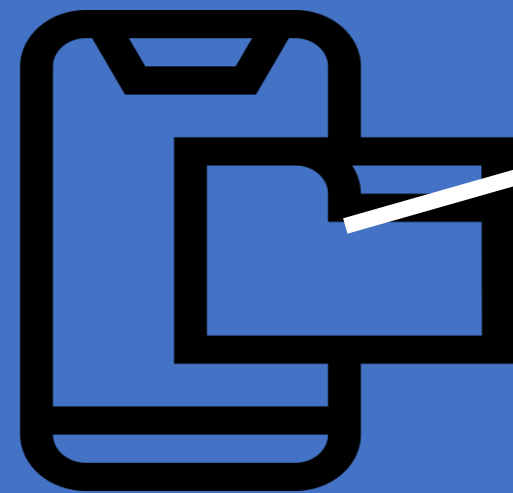
활용방안 #1



수강과목 추천

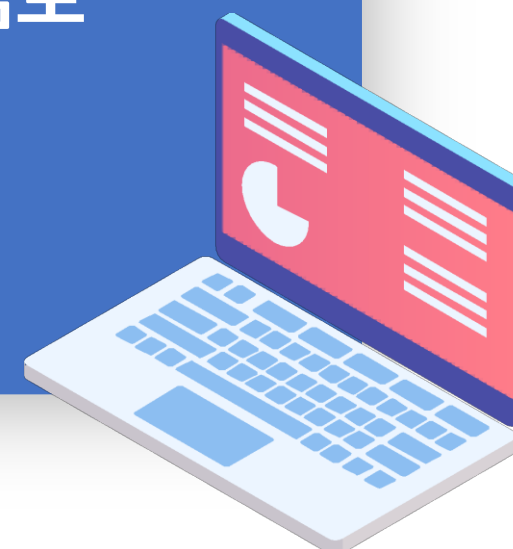
증명서 발급 서비스를 통한 성적증명서에
‘다음학기 수강과목추천’ 항목을 추가해
학생들이 실질적인 도움을 받을 수 있도록
활용

활용방안 #2



수강과목 추천

각종 커뮤니티 어플을 통해 손쉽게
추천과목을 확인할 수 있도록
서비스를 추가하는 방향 검토



기대효과

인적네트워크의 의존도 높았던
수강과목 선정을 AI가 대신 해주어
수강의 효율성을 개선

직무관련도가 높은 수강과목
추천으로 취업에 도움



참고) 디지털존 멘토링 계획서

4.3. 한계점 및 개선점

한계점

- 01** 다양한 범주값에 비해 적은 데이터 수로, 여러 모델을 시도하기에는 어려움이 있었음
- 02** Auto ML 을 사용해 자연어 처리를 시도했으나 다양한 범주의 값으로 인해 제대로 분류되지 않았음
- 03** 1987년도부터 수집된 데이터, 특정 직업에 편중된 데이터로 현 시대의 트렌드를 반영하지 못한 아쉬움이 있었음
- 04** 정확도, F-1 스코어를 높이기 위해 학생들이 많이 들은 과목을 위주로 추천해주는 모델을 만들어 다양한 과목 추천의 어려움이 있었음

개선점

- 01** 프로젝트 진행초반 논문 등과 같은 선행연구에 관한 정보를 통해 도움을 얻을 수 있도록 함
- 02** 학교별 모델을 만들어 정확도가 높은 학교별 맞춤 서비스 필요
- 03** 학생들의 강의평가와 같은 여러 정보를 추가해 수강신청시 다방면으로 도움을 받을 수 있도록 발전
- 04** 여러 서비스가 융합된 플랫폼을 개발하여 활용

Thank you

Q&A