

맞춤형 팀플 및 과제 부담 최소화 강의 분류 시스템

25-2 기계학습의 이해 프로젝트 최종보고서

202401394 박의진

목차

- 프로젝트 개요
- 진행 과정
- 모델을 서비스로 만든 구조
- 실제 사용 결과
- 배운 점 및 개선 방향

1. 프로젝트 개요

1-1 무엇을 만들었나?

맞춤형 팀플 및 과제 부담 최소화 강의 분류 시스템은 한국외국어대학교 강의 계획서의 텍스트 데이터를 분석하여, 해당 강의의 팀플(조별 과제, 프로젝트 등) 유무와 과제 부담 지수 생성하여 이를 예측해 주는 AI 웹 서비스입니다. 사용자가 강의계획서 전체를 복사하여 붙여넣기만 하면, AI가 분류에 필요한 핵심 텍스트를 추출하고 분석 결과를 통해 예측합니다.

1-2 어떤 문제를 해결했나?

한국외대 대학생들의 수강신청 기간 최대 고민 중 하나는 강의평과 대학생 커뮤니티인 에브리타임을 보면 "이 수업에 팀플이 있는가?"와 "과제가 많은가?"입니다. 강의계획서에는 '팀 프로젝트', '조별 발표', '조별 과제' 등 다양한 표현이 혼재되어 있고, 각각 명시되어 있는 위치 또한 교수님마다 매우 다르기 때문에 텍스트를 일일이 읽어보지 않으면 파악하기 어렵습니다. 제 프로젝트는 이러한 비정형 텍스트를 머신러닝 모델로 분석하여, 팀플을 꺼려하고 과제가 적은 소위 말하는 "꿀강"을 원하는 대학생에게 확률을 높이고 학생들의 대학 삶의 질을 높이는 것을 목표로 합니다.

2. 진행과정

2-1 주제 선정 및 문제 정의

주제 : 강의계획서 텍스트를 통한 팀플 유무/과제 부담도 탐지기

1. 팀플 유무 분류 : Binary Classification Task
2. 과제 부담도 예측 : Regression Task

2-2 데이터 수집 및 분석

데이터 : 2025학년도 1,2학기 한국외국어대학교 강의계획서 100건 수집

Preprocessing :

1. Regex를 활용하여 교과목명, 수업개요, 수업운영방식, 기타 안내의 교수님들께서 많이 작성하시는 4가지 영역만 추출 및 하나로 병합
2. 기타 점수(EtcScore), 키워드('조별','팀' 등) 기반 라벨링
3. Train/Validation/Test 8:1:1로 나누고 시작

2-3 ML 모델 학습 및 평가

모델 선정 : 한국어 텍스트 이해도가 높고 추론 속도가 빠른 모델이 필요하지만 가지고 있는 환경에서 빠르게 돌아가기 위해 경량화가 필요하다고 판단 → monologg/distilkobert 사용

학습 : 정확도를 높이기 위해 팀플 분류 모델과 과제 부담 회귀 모델을 각각 독립적으로 학습시킴.

평가 결과 :

Recall : 0.83 (실제 팀플 수업을 83% 확률로 탐지)

Accuracy: 0.60 (데이터 부족으로 인한 일부 오탐지 존재).

2-4 서비스 개발

Framework : Python Streamlit으로 웹 애플리케이션 구축.

기능 구현 :

1. 강의계획서 전체 텍스트 자동 파싱 기능 구현
2. 팀플 위험도 및 과제 부담 지수 표현
3. AI 판단 근거(키워드) 제공 및 사용자 피드백 시스템 구축.

3. 모델을 서비스로 만든 구조

3-1 시스템구조

학습된 모델을 Google Drive에 호스팅, Streamlit Cloud 서버가 이를 다운로드하여 추론하는 구조로 제작해 사용자는 별도의 다운없이 바로 링크로 접속하여 모델을 사용할 수 있게 하였습니다.

1. User Interface: 사용자가 웹 브라우저에서 강의계획서 텍스트를 복사 붙여넣기로 입력.
2. Preprocessing Engine: Python regex 모듈이 입력된 텍스트에서 불필요한 정보(날짜, 학수번호 등)를 제거하고 핵심 문단만 추출.
3. Model Inference:
 - DistilKoBERT Tokenizer가 텍스트를 토큰화.

- Model A (Classifier)가 팀플 확률(%) 예측.
- Model B (Regressor)가 과제 부담 점수 예측.
- 4. Post-processing: 팀플 확률이 높을 경우 과제 부담 점수에 가중치(1.5배)를 부여하여 최종 '기피강의/꿀강' 등급 산정.
- 5. Visualization: 결과 및 판단 근거 텍스트를 UI에 표시.

About Project

AI가 강의계획서를 통해 팀플 유무와, 과제 부담 지수를 예측하여 분류합니다.

Student ID/Name:
202401394 박의진

Model: Dual DistilKoBERT

Credit: Gemini pro 3의 도움을 받아 제작하였습니다.

팀플 및 과제 부담 최소화 강의 분류 시스템

강의계획서 전체를 복사해서 붙여넣으세요!

AI가 분석에 필요한 텍스트만 추출하여 분석하겠습니다.

강의계획서 전체 텍스트 붙여넣기.

종합정보시스템 강의계획서 화면을 전체 복사(Ctrl+A, Ctrl+C)해서 여기에 붙여넣으세요.

강의 분류 시작

About Project

AI가 분석에 필요한 텍스트만 추출하여 분석하겠습니다.

강의계획서 전체 텍스트 붙여넣기:

(4) 대인관계 역량(Interpersonal Competency)
 - 정서적 유대 및 공감 능력(Emotional ties and empathy)

(5) 세계시민 역량(Global Civic Competency)
 - 문화적 감수성(Cultural sensitivity)

(6) 의사소통 역량(Communication Competency)
 - 연관성 있음(N/A)

핵심영역 추진 전략
 (Strategy for Core Competency-Within1000 characters)
 25. 11. 26. 오후 3:17 wis.hufs.ac.kr/src08/jsp/lecture/syllabus.jsp?
 mode=print&ledg_year=2025&ledg_sessn=3&org_sect=A&lssn_cd=U51108101
[https://wis.hufs.ac.kr/src08/jsp/lecture/syllabus.jsp?](https://wis.hufs.ac.kr/src08/jsp/lecture/syllabus.jsp?mode=print&ledg_year=2025&ledg_sessn=3&org_sect=A&lssn_cd=U51108101)
 mode=print&ledg_year=2025&ledg_sessn=3&org_sect=A&lssn_cd=U51108101 4/4

강의 분류 시작

분석 결과: 신화와 종교 (Myth & Religion)

팀플 위험도	과제 부담 지수
 팀플 있음 (61.9%)	Burden Score 50.3/100점
조별 과제가 감지되었습니다. 수강 시 유의하세요!	
보통	

4. 실제 사용 결과

1) 기계학습의 이해(전공선택)

📁 분석 결과: 기계학습의 이해 (UnderstandingMachine Learning)

팀플 위험도	과제 부담 지수
 팀플 있음 (63.6%)	Burden Score
조별 과제가 감지되었습니다. 수강 시 유의하세요!	50.3/100점

➤ ◎AI가 출출한 핵심 내용 보기 (판단 근거)

▶ 분석 결과가 정확한가요?

 정확해요  틀렸어요

 다른 강의 분석하기

팀풀이 없으나, 팀풀 있음으로 분류 실패, 과제 부담도는 체감 상 50.3보다 높다고 생각하지만, 이는 계획서 텍스트로는 파악 한계가 있음.

2) 윤리적 사고와 문제 해결(교양)

https://wis.huts.ac.kr/src08/jsp/lecture/syllabus.jsp?mode=print&legd_year=2025&legd_sessn=3&org_sect=A&lssn_cd=Y211891015/5

강의 분류 시작

📁 분석 결과: 윤리적 사고와 문제 해결 (Moral Issues and Resolution)

팀플 위험도

과제 부담 지수

⭐ 팀플 있음 (75.9%)

조별 과제가 감지되었습니다. 수강 시 유의하세요.

Burden Score

52.0/100점

보통

> ◎AI가 추출한 핵심 내용 보기 (판단 근거)

 분석 결과가 정확한가요?

 정확해요

팀 프로젝트가 존재하는 수업. 팀플 있음으로 높은 확률 분류 성공, 시험/토론 위주라 과제지수는 52점 예측

3) 동유럽.발칸문화의이해(교양)

📁 분석 결과: 동유럽.발칸문화의이해 (Cultureof Eastern Europe & Balkans)

팀플 위험도

팀플 없음 (39.4%)

개인 과제 위주이거나 이론 수업일 확률이 높습니다.

과제 부담 지수

Burden Score

8.8/100점

꿀강 확정 🍎 (과제 적음)

> ⓘ AI가 추출한 핵심 내용 보기 (판단 근거)

💡 분석 결과가 정확한가요?

👍 정확해요

👎 틀렸어요

과제물 0%로, 과제 부담 지수 매우 낮게 예측 성공

전체 시험 90%로 평가인 과목이기에, 팀플 없음으로 예측 성공

4) 현대인의 정보 관리

📁 분석 결과: 현대인의정보관리 (InformationManagement of the Moderns)

팀플 위험도

팀플 없음 (32.9%)

개인 과제 위주이거나 이론 수업일 확률이 높습니다.

과제 부담 지수

Burden Score
24.1/100점

꿀강 확정 🍎 (과제 적음)

> ⓘ AI가 추출한 핵심 내용 보기 (판단 근거)

💡 분석 결과가 정확한가요?

👍 정확해요

👎 틀렸어요

➡ 다른 강의 분석하기

팀플 없음 분류 성공, 과제 부담 지수 낮게 예측.

5) 한중일문학산책

📁 분석 결과: 한중일문학산책 (Cultural Excursions: Korea, China, and Japan)

팀플 위험도	과제 부담 지수
⚠️ 팀플 있음 (80.8%)	Burden Score 53.0/100점
조별 과제가 강지되었습니다. 수강 시 유의하세요!	보통

➤ AI가 주출한 핵심 내용 보기 (판단 근거)

📌 분석 결과가 정확한가요?

🔥 정확해요 🌟 둘렀어요

➡ 다른 강의 분석하기

"팀플 발표는 시행하지 않을 예정이다"라는 텍스트를 "팀플"이라는 텍스트만 탐지하여 팀플 위험도를 높게 예측. 팀플 있음으로 분류 실패

실제 사용 후 "동유럽·발칸문화의 이해" 과목이 꿀강이라는 것을 에브리타임 및 주위 학우들 의견 교차검증을 통해 확인. 실제 다음 학년 교양 수업으로 수강하기로 결정. 또한 현대인의 정보관리 강의 또한 수강 예정

5. 배운 점 및 개선 방향

모델을 가져와 학습시키고 원하는 task를 진행하는 것은 다른 전공 과제에서도 진행해 보았으나, 이를 웹 서비스로 제작하여 Demo를 배포하는 것은 경험이 적기에, 이번 프로젝트를 통해 Demo 구현 파이프라인까지 최종적으로 경험하였습니다. 또한 모델만이 중요한 게 아니라, task를 위한 데이터셋이 정말 중요하다는 것을 다시 한번 경험했습니다. 제가 사용한 데이터셋 100개도 모으는 것은 쉽지 않은 양이지만, 모델이 이 데이터셋을 가지고 완벽하고 정확하게 분류하기엔 매우 적은 양이었습니다. 100개가 아닌 1000개였다면 아마 성능이 훨씬 좋았을 것이라 생각합니다. 인터페이스를 구현하는 과정이 저에게 정말 흥미로운 과정이었습니다. 그저 버튼만 누르면 돌아가게 구현하는 것이 아닌, 사용자에게 서비스한다는 생각으로 예쁜 디자인은 아니지만 사용자 편의를 고려하며 제작했던 것 같습니다.

개선에 관련한 방향은 이렇습니다. 우선 데이터가 매우 많이 필요할 것 같습니다. 위의 5번 실제 사용 후기를 통해도 보셨듯이, 팀플 유무에 대하여 정확히 분류하지 못하는 성능을 드러냈고, 이는 근본적으로 데이터셋이 부족하다고 생각했습니다. 따라서 가장 기초적인 데이터셋 수집. 두 번째로는 데이터 라벨링 과정입니다. 사용 후기 5번처럼 "팀플은 없음"이라는 텍스트에서 현재 모델은 "팀플"이라는 텍스트에만 집중하여 팀플을 있음으로 분류하는 경향이 드러났습니다. 이를 방지하기 위해 라벨링 과정에서 "팀플"이라는

워딩에만 집중하지 않고, '없음'이나 '지양'과 같은 부정 표현이 인접해 있는지를 함께 검토하는 '부정어 필터링' 로직을 추가할 필요를 사용하며 느꼈습니다. 마지막으로 현재는 사용자가 텍스트를 전체 복사하여 붙여넣는 방식이지만, 추후 개선한다면 Bot 형태로 강의계획서 창을 열면, 텍스트를 복사하지 않아도 화면을 통해 탐지하여 바로 분류해주는 구조로 발전시키고 싶습니다. 그러면 정말로 계획서를 켜서 바로바로 분류하여 수강신청을 매우 편리하게 할 수 있는 구조가 될 것입니다.