

# 프로젝트 최종 보고서

과목명: 기계학습의 이해

학번 이름: 2020848 조현아

주제: 학교 공지 통합 알림 봇

## 1. 프로젝트 개요

### 1.1. 무엇을 만들었나

프로그램을 실행하면 학교 공지사항을 자동으로 크롤링하여 사용자(저)에게 관련 있는 공지사항만 중요도에 따라 알림 방식을 다르게 하는 프로그램을 제작했습니다. 학교 메인 홈페이지의 일반 공지, 학사, 장학, 국제통상학과(이하 IEL) notice와 자유게시판, Language & AI 융합학부(이하 LAI)의 공지사항과 학과 소식 페이지를 크롤링, 학습 결과에 따라 노선과 이메일의 알림 방식을 달리했습니다.

### 1.2. 어떤 문제를 해결했나

학교 공지사항의 종류에는 여러 가지가 있습니다. 졸업에 관련된 것부터 각종 특강이나 행사, 장학, 취업에 도움되는 소식 등 여러 내용이 여러 개의 페이지에 분산되어 올라옵니다. 매번 다른 페이지를 열어서 내용을 모두 읽어보는 것은 꽤 시간이 소요되는 작업이고, 규칙을 정하고 들어가는 것이 아니기에 이미 읽었던 공지사항을 또 읽는 상황도 빈번합니다. 저와 관련 있는 공지만 시간낭비 없이 확인하고, 중요한 공지사항은 놓치지 않도록 했습니다.

## 2. 진행 과정

### 2.1. 주제 선정 및 문제 정의 (Assignment 3)

학교생활을 하며 머신러닝을 적용해서 제 불편함을 해소할 수 있는 방안에 대해 생각하고 주제를 결정했습니다.

## 2.2. 데이터 수집 및 분석 (Assignment 4)

### 2.2.1. 데이터 수집

- 학교 메인 홈페이지의 일반공지, 학사, 장학, IEL 공지와 자유게시판, LAI 공지사항과 학과소식을 52페이지까지 크롤링해서 완전히 중복되는 공지사항은 1개만 남기고 제거했습니다.
- 관련도 0,1, 중요도 0,1,2로 라벨링한 뒤 주기적으로 반복되는 공지를 제거했습니다(예시. 매달 반복되는 심폐소생술 교육 관련 공지를 제거했습니다).
- 추가 전처리로 연도, 전화번호, "안녕하세요. Language & AI 사무실입니다."를 제거했습니다. 연도를 제거한 이유는 중복되는 공지를 제거하다 보니 주로 2025년 공지가 남아서, 2026년 등 다른 연도가 입력되었을 때 자동으로 관련이 없다고 처리하는 것을 방지하기 위해, 전화번호를 제거한 이유는 학습을 하지 않고 학교 관련 부서의 전화번호만 보고 관련도나 중요도를 판단해 버리는 상황을 방지하기 위해서입니다. 예를 들어, 글로벌캠퍼스 부서의 공지여도 서울캠퍼스 학생들도 대상인 공지 내용을 글로벌캠퍼스 부서의 전화번호가 있다고 내용을 보지 않고 자동으로 관련도 0으로 처리하는 등의 상황을 방지하고자 했습니다. "안녕하세요. Language & AI 사무실입니다."를 제거한 이유는 상대적으로 LAI의 공지에 중요도1, IEL의 공지에 중요도 0인 공지가 많아서였습니다. IEL같은 경우 이런 고정적으로 자주 등장하는 문구가 없어 그대로 두었고, LAI는 이 문구가 등장하는 경우 관련도 1이었던 경우가 많아 제거했습니다.

### 2.2.2. 데이터 분석

- 데이터셋에서 불균형이 나타났습니다. 관련도는 0:1의 비율이 약 2:3으로 상대적으로 제 데이터 안에서는 균형이 있는 편이었지만 중요도는 편차가 심했습니다(중요도 0:1:2 = 948:284:123). 출처별 공지의 경우 52페이지씩 크롤링한 뒤 반복되는 공지를 제거하고 나니 학사 공지는 거의 2/3이 제거되었습니다.

## 2.3. ML 모델 학습 및 평가 (Assignment 5)

- 학습에는 관련도, 중요도, 제목, 본문만 사용했습니다. 데이터셋에는 출처도 표시되어 있으니 이는 성능 평가를 위해 넣은 부분이고 학습 단계에서는 무시했습니다.
- 모델은 Klue RoBERTa Small을 사용했습니다. RoBERT의 한국어 특화 모델로, 데이터셋 개수가 적어 Small 모델을 사용했습니다.

- train : validation : test = 8 : 1 : 1
- 각 출처별로 시계열로 분할하여 가장 최신 데이터가 test dataset가 되도록 하였습니다. 출처별로 분할했기 때문에 각 데이터셋에서 출처의 비율은 일정하게 유지했지만 관련도나 중요도의 비율은 일정하지 않았습니다.
- evaluation method: Macro F1-score를 주된 방법론으로 사용했습니다. Macro를 사용한 이유는 중요도 2,1,0 순으로 맞추는 것이 중요하지만 데이터셋의 개수는 0,1,2순으로 많아서입니다.
- 데이터 증강: 데이터 증강 이전 IEL, 장학의 공지사항들이 다른 출처보다 성능이 확연히 낮았습니다. 문장 길이의 15%만큼 랜덤한 문장부호를 삽입하는 방식으로 데이터 증강을 진행했고, IEL, 장학은 3개씩, 나머지는 1개씩 증강을 진행했습니다.
- 다른 하이퍼파라미터들도 조절해 가며 F1-score를 높여 나갔습니다.

#### 2.4. 서비스 개발 (Assignment 6)

- 평소에 자주 사용하는 노선에 공지사항을 정리하는 페이지를 만들고 중요도2, '선착순'이라는 단어가 있는 공지는 이메일을 발송하도록 제작했습니다.
- 노선에는 공지의 제목과 제가 라벨링한 중요도, 공지 날짜, 그 공지사항의 링크를 넣어 두었고, 이메일은 중요한 것만 보냈으니 본문까지 포함해서 보내도록 설계했습니다.
- 노선 자동 등록과 이메일 자동 발송 방식을 사용했습니다. 관련도가 1인 것은 노선에 등록되고, 그 중 중요도가 1,2인 것은 달력에도 등록되게 하였습니다. 중요도가 2인 것과 '선착순'이라는 단어가 들어간 것은 이메일이 발송되도록 만들었습니다.

### 3. 모델을 서비스로 만든 구조

#### 3.1. 시스템 구조 (API, CLI, 봇 등)

- API: 노선 데이터베이스 ID와 API로 노선에 자동 등록되게 만들었고, GPT 4o-mini를 사용해 공지사항 본문 안에 있는 공지와 마감일을 추출하도록 했습니다.
- Training은 GPU로 진행했지만 실사용은 CPU로 진행하기 때문에 assignment5의 모델 가중치를 model/ 폴더 안에 저장해야 합니다.
- history.json: 크롤링한 공지사항을 저장해 두는 파일로 이미 분류한 공지사항을 또 분류하는 것을 방지합니다.
- crawler.py: 지정한 홈페이지의 1페이지만 크롤링합니다. history.json에 없는 공지

- execution\_log.txt: 과제 제출용으로 만든 프로그램 로그 기록입니다.

## School Notice

**School Notice**

제목	작성일	종류
< 삼성생명X한국국립대학교 협찬스 티즈 >		Black
[LA Event Trading Corp] 멀티버전/그라운드		Black
2025 연세호팅 H&F도감		Black
[2025 H&F 품목 인도식대매]		Black
삼성 청년SW-AI아카데미(SSAFY) 14기 모집	2025년 12월 4일	Red
2025년도 우리은행 학제 제정당 인건 제공 (보)		Red
[백양수신부] 2025 SICE IMO	2025년 10월 22일	Red
[미래창조과학연구원] 2025 PECC 유망 학력		Red
[D&S에너지] 신입사원 공개채용 최종 상담부		Black
[Empower Worl] Miria Danner Abreu 주		Black
2025년도 2학기 1차 졸업시험 결과 안내		Black
2025년도 2학기 2차 졸업시험 일정 공지 및		Black

## School Notice

날짜 없음(69) 새로 만들기

2025년 12월

캘린더에서 관리하기

일	월	화	수	목	금	토
30	12월 1일	2	3	4	5	6

- 삼성 청년SW-AI아카데미(SSAFY) 14기 모집
- [일반공지 ...]
- [모두가 주...]
- Language...
- AI 융합데...
- [기타] [메...]
- [기타] 26...
- [7+1/국외...]
- [공동] 202...

---

## School Notice

**School Notice**

제목	작성일	종류
< 삼성생명X한국국립대학교 협찬스 티즈 >		Black
[LA Event Trading Corp] 멀티버전/그라운드		Black
2025 연세호팅 H&F도감		Black
[2025 H&F 품목 인도식대매]		Black
삼성 청년SW-AI아카데미(SSAFY) 14기 모집	2025년 12월 4일	Red
2025년도 우리은행 학제 제정당 인건 제공 (보)		Red
[백양수신부] 2025 SICE IMO	2025년 10월 22일	Red
[미래창조과학연구원] 2025 PECC 유망 학력		Red
[D&S에너지] 신입사원 공개채용 최종 상담부		Black
[Empower Worl] Miria Danner Abreu 주		Black
2025년도 2학기 1차 졸업시험 결과 안내		Black
2025년도 2학기 2차 졸업시험 일정 공지 및		Black

## School Notice

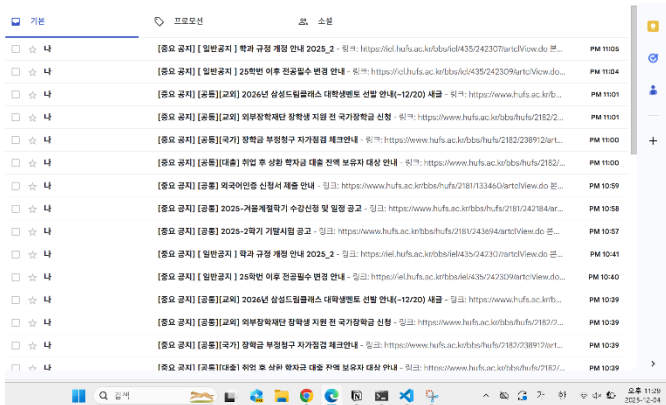
날짜 없음(69) 새로 만들기

2025년 12월

캘린더에서 관리하기

일	월	화	수	목	금	토
30	12월 1일	2	3	4	5	6

- 삼성 청년SW-AI아카데미(SSAFY) 14기 모  
집 안내
- [일반공지 ...]
- [모두가 주...]
- Language...
- AI 융합데...
- [기타] [메...]
- [기타] 26...
- [7+1/국외...]
- [공동] 202...



[중요 공지] [일반공지] 학과 규정 개정 안내 2025\_2

a

링크: <https://iel.hufs.ac.kr/bbs/iel/435/242307/andView.do>

본문:

학과 규정 개정 안내 2025\_2 \* 학과 교수진 회의를 통해 학과 규정이 개정되었습니다. 새 규정을 첨부합니다. (바뀐 내용은 붉은 색 처리) \* 개정된 내용은 다음과 같습니다.

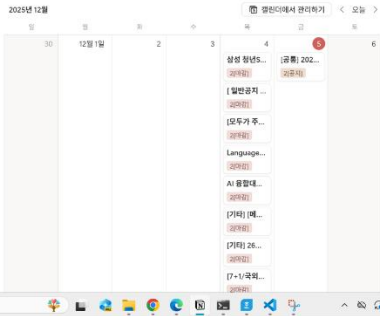
답장 전발

## 4.1.2. 12월 5일 금요일

수정 사항: 이메일을 보내는 경우 공지 한 개 당 이메일 한 개가 발송되는 것 대신 발송해야 할 공지를 한번에 발송하도록 수정했습니다. 중요도 2인 경우 2[공지], 2[마감]으로 분류해서 공지일과 마감일 모두가 노선 달력에 표시되도록 했습니다. API를 Gemini 1.5 flash에서 GPT-4o-mini로 변경했습니다.

### Calendar

#### School Notice



### School Notice

#### Calendar

#### School Notice

Aa 제목	링크	마감일	중요도
[공통][국가균등] 2025년 2학기 국가균등	<a href="https://iel.hufs.ac.kr/bbs/iel/435/242307/andView.do">hufs.ac.kr/bbs/iel/435/242307/andView.do</a>	2025년 12월 8일	1
[서울][국가균등] 2025학년도 2학기 국가균	<a href="https://iel.hufs.ac.kr/bbs/iel/435/242307/andView.do">hufs.ac.kr/bbs/iel/435/242307/andView.do</a>	2025년 12월 8일	1
[공통] 2025-거울계절학기 등록기간 및 환불	<a href="https://iel.hufs.ac.kr/bbs/iel/435/242307/andView.do">hufs.ac.kr/bbs/iel/435/242307/andView.do</a>	2025년 12월 8일	2[마감]
[공통] 2025-거울계절학기 등록기간 및 환불	<a href="https://iel.hufs.ac.kr/bbs/iel/435/242307/andView.do">hufs.ac.kr/bbs/iel/435/242307/andView.do</a>	2025년 12월 5일	2[공지]
[공통] 2025학년도 제1학기 재입학 시험 공	<a href="https://iel.hufs.ac.kr/bbs/iel/435/242307/andView.do">hufs.ac.kr/bbs/iel/435/242307/andView.do</a>		0
2025학년도 동계방학 「Hufs 사회봉사단」	<a href="https://iel.hufs.ac.kr/bbs/iel/435/242307/andView.do">hufs.ac.kr/bbs/iel/435/242307/andView.do</a>		0
<상상생명>한국의국립대학교 캠퍼스 리크	<a href="https://iel.hufs.ac.kr/bbs/iel/435/242307/andView.do">iel.hufs.ac.kr/bbs/iel/435/242307/andView.do</a>		0
[LA Everest Trading Corp] 팍다자/그라	<a href="https://iel.hufs.ac.kr/bbs/iel/435/242307/andView.do">iel.hufs.ac.kr/bbs/iel/435/242307/andView.do</a>		0
2025 명사초청 Hufs특강	<a href="https://iel.hufs.ac.kr/bbs/iel/435/242307/andView.do">iel.hufs.ac.kr/bbs/iel/435/242307/andView.do</a>		0
[2025 Hufs] 동문 연회동대	<a href="https://iel.hufs.ac.kr/bbs/iel/435/242307/andView.do">iel.hufs.ac.kr/bbs/iel/435/242307/andView.do</a>		0
상상 청년SW AI아카데미(SSAFY) 14기 모	<a href="https://iel.hufs.ac.kr/bbs/iel/435/242307/andView.do">iel.hufs.ac.kr/bbs/iel/435/242307/andView.do</a>	2025년 12월 4일	2[마감]
2025년 우리은행 학계 제정할 만한 재	<a href="https://iel.hufs.ac.kr/bbs/iel/435/242307/andView.do">iel.hufs.ac.kr/bbs/iel/435/242307/andView.do</a>		0

[학교 공지 알림] 중요 공지 1건 도착

a

나에게  
총 1건의 중요 공지가 있습니다.

[중요] [공통] 2025-거울계절학기 등록기간 및 환불 안내

링크: <https://www.hufs.ac.kr/bbs/hufs/2181/244062/andView.do>

내용 미리보기:

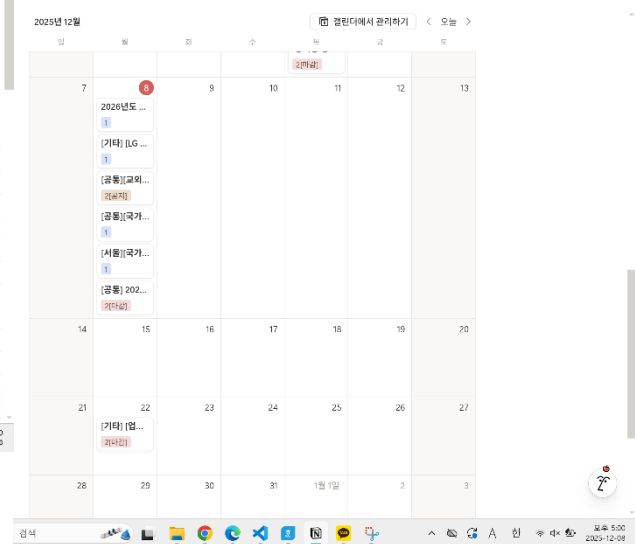
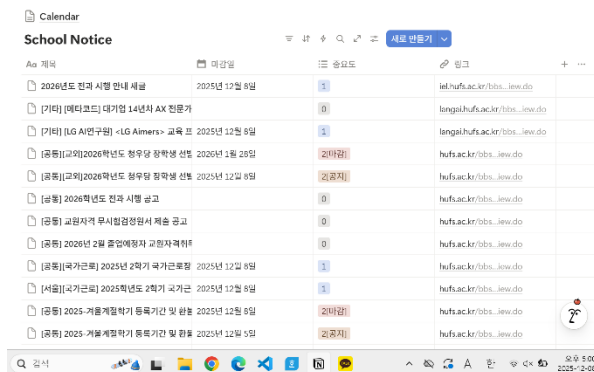
2025-거울계절학기 등록기간 및 환불 안내 1. 수강료 납부 및 고지서 출력 등록기간 2025. 12. 8.(월) ~ 12. 10.(수) ※ 수강신청자에 한하여 납부 가능함 (미수강신청자는 납부 불가) 2. 수강료 학점당 85,000원 수강료 확인 (고지서 출력) \* 계절학기 등록금 고지서 출력(여-출력) \* (홈페이지 핫링크 또는 종합정보시스템) 납부방법 - 우리은행 전국지점 납부(영업시간 내) - 고지서에 있는 본인명의 우리은행 가상계좌에 입금 (가상계좌 이용시간: 08:00~16:00) 납부확인 (발입 파일) ① 실시간 : 우리...

답장 전발

#### 4.1.3. 12월 8일 월요일

수정 사항: 중요도 상관 없이 '선착순'이라는 단어가 있으면 메일 발송되도록 수정했습니다. 기존에는 중요도 2인 경우에만 발송되었습니다. 공지일과 마감일을 잘 맞추지 못해서 Gpt-4o-mini의 프롬프트를 영어에서 한글로 수정했습니다(내용은 동일, 번역만 수행).

#### School Notice



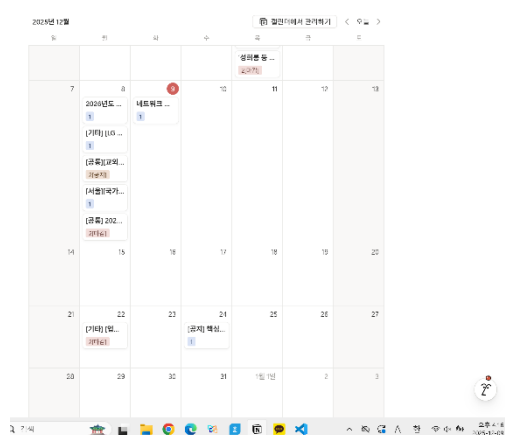
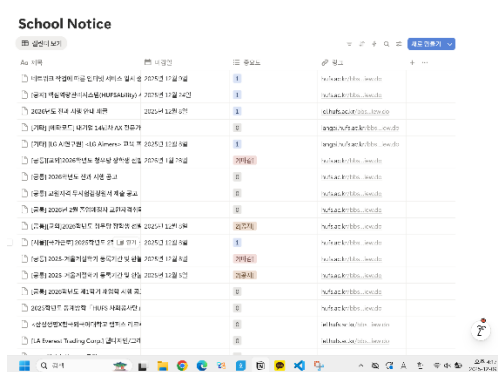
[학교 공지 알림] 중요 공지 1건 도착 받은편지함 x

나에게  
총 1건의 중요 공지가 있습니다.

[중요] [공통][교외]2026학년도 청우당 장학생 선발 안내(~1/28)  
링크: <https://www.hufs.ac.kr/bbs/hufs/2182/244144/articleView.do>  
내용 미리보기:  
2026년 청우당 장학금 장학생 선발 공고 1. 청우당(淸雨堂) 장학금 개요 2. 장학목적: 여름방학 중 해외연수 지원 장학금(등록금 및 장학금-생활비상) 나. 수여식 일정: 2026년 3월 6일(금). 15:00, Onycom(주) 용산사옥 5층 다. 해외연수보고서 제출: 2026년 9월 중 (PDF 파일 1개, 텍스트(wordhwp)만 있는 파일 1개, 이미지 파일을 각각 첨부하여 [alumni@hufs.ac.kr](mailto:alumni@hufs.ac.kr)로 제출 2. 장학생 선발 가. 신청자격: 양 캠퍼스 2~4학년 재학생 ※ 반드시 2026년 1~2학기 연속으로 재학...

← 답장 → 전달

#### 4.1.4. 12월 9일 화요일



\* 이메일 발송 대상 공지 없음.

#### 4.1.5. 12월 10일 수요일

## School Notice

Ao	제목	마감일	종료일	정기	+ ...
[구동]	2025-거울개발팀이 동국대 연명 >	2025년 12월 21일	2<대강>	huft.ac.kr/bbs_iew.do	
[공회]	2025-거울개발팀이 동국대 연명 >	2025년 12월 10일	2<교지>	huft.ac.kr/bbs_iew.do	
2026학년도 1학기 코트라(KOTRA) 해외무 >	2025년 12월 16일	2<연강>	huft.ac.kr/bbs_iew.do		
2026학년도 1학기 코트라(KOTRA) 해외무 >	2025년 12월 10일	2<기술>	huft.ac.kr/bbs_iew.do		
내세부 작업에 따른 인터넷 서비스 일시 >	2025년 12월 15일	1	huft.ac.kr/bbs_iew.do		
[삼지] 특성화평판관리시스템(HUFAbility) >	2025년 12월 24일	1	huft.ac.kr/bbs_iew.do		
2026년도 전과 시험 안내 새글	2025년 12월 8일	1	ie1.huft.ac.kr_bbs_iew.do		
[가리][태바코즈] 대입실 14선 AX 선문 >		0	lang1.huft.ac.kr/bbs_iew.do		
[가리][LG AI연구관] <LG Almers - 교육 >	2025년 12월 4일	1	huft.ac.kr/bbs_iew.do		
[공회]고교2026학년도 원후보 작성법 생애 >	2025년 1월 28일	2<대강>	huft.ac.kr/bbs_iew.do		
[공회] 2026학년도 전과 시험 공고		0	huft.ac.kr/bbs_iew.do		
[공회] 교본자적 과목집합원서 제출 공고		0	huft.ac.kr/bbs_iew.do		

2025년 12월

월 화 수 목 금 토

7	8	9	10	11	12	13
2020년도 ... [기] 1 [기타] [LG ... [기] 1 [공통] 교육 ... [공과자] [서울] 국가 ... [기] 1 [공통] 202... [과제자]	네트웍스 ... [기] 1	[공통] 202... [과제자] 2026학년... [과제자]	[과제자]			
14	15	16	17	18	19	20
	2026학년... [과제자]					
21	22	23	24	25	26	27
[공통] 202... [과제자]	[기타] [연... [과제자]	[공과자] 박성... [기] 1				
28	29	30	31	1월 1일	2	3

2025.12.15

[학교 공지 알림] 중요 공지 2건 도착 받은편지함 x

나에게

총 2건의 중요 공지가 있습니다.

=====

[중요] 2026학년도 1학기 코트라(KOTRA) 해외무역관 현장실습 모집 안내(-12.16, 17시) 새글  
 링크: <https://www.hufs.ac.kr/bbs/hufs/2180/244318/artclView.do>  
 내용 미리보기:  
 2026학년도 1학기 코트라(KOTRA) 해외무역관 현장실습 모집 안내(-12.16, 17시) 추천후계 게시판에서 온라인 지원 2026학년도 1학기 KOTRA(Korea Trade-Investment Promotion Agency, 대한무역투자진흥공사) 해외무역관 현장실습 모집일정을 아래와 같이 안내드리오니 기한 내 공지내용을 준수하여 지원 바랍니다. - 아 래 - 1. 파견학기 : 2026학년도 1학기 2. 선발인원 : 각 무역관 별 명額 [2026년 1학기 파견 가능 무역관] (가나다순) 순번 무역관명 순번 무역관명 1 나고라 ...

=====

[중요] [공통] 2025-겨울계절학기 등록기간 연장 및 환불 안내  
 링크: <https://www.hufs.ac.kr/bbs/hufs/2181/244344/artclView.do>  
 내용 미리보기:  
 2025-겨울계절학기 등록기간 연장 및 환불 안내 2025학년도 겨울계절학기 수강료 납부 마감일을 아래와 같이 연장하오니 수강신청 후 수강료를 납부하지 않은 학생은 수강료 납부를 완료하시기 바랍니다. 1. 수강료 납부 및 교지서 출력 등록기간 12. 11.(목) ~ 12. 12.(금) ※ 수강신청처지에 한하여 납부 가능함 (미수강신청자는 납부 불가) ※ 강 료 학점당 85,000원 수강료 확인 (교지서 출력) 계절학기 등록금 교지서 출력(홈페이지 하위공지(마포가나) 또는 종합정보시스템(등록/전정정보) > 계절학기 교지서발급) 납부...

#### 4.1.6. 12월 11일 목요일

😊 아이콘 추가   🖼️ 커버 추가   💬 댓글 추가

## School Notice

Calendar

School Notice

새로 만들기

Aa 제목	📅 마감일	⋮ 중요도	🔗 링크
📄 [공통][교외] DB김준기문화재단 2026년 제...		0	<a href="https://hufs.ac.kr/bbs...iew.do">hufs.ac.kr/bbs...iew.do</a>
📄 [서울][교외] 2026학년도 성육문화재단 장...		0	<a href="https://hufs.ac.kr/bbs...iew.do">hufs.ac.kr/bbs...iew.do</a>
📄 [공통][교외] 2026학년도 전기 양천장학회...		0	<a href="https://hufs.ac.kr/bbs...iew.do">hufs.ac.kr/bbs...iew.do</a>
📄 [글로벌][국가근로] 2025학년도 2학기 동...		0	<a href="https://hufs.ac.kr/bbs...iew.do">hufs.ac.kr/bbs...iew.do</a>
📄 [글로벌][국가근로] 2025학년도 2학기 동...		0	<a href="https://hufs.ac.kr/bbs...iew.do">hufs.ac.kr/bbs...iew.do</a>

\* 전부 중요도가 0이라 달력에 추가될 공지와 발송할 이메일 모두 없음

## 5. 배운 점 및 개선 방향

실제 사용 결과 관련도는 잘 맞췄지만 중요도2인 공지나 장학 카테고리의 공지사항에 대한 성능이 상대적으로 가장 좋지 않았습니다. 장학의 경우 제게 맞는 조건의 복잡도에 비해 데이터셋의 개수가 적었고, 중요도2 역시 데이터셋의 개수가 적어서 발생한 결과라 생각됩니다. 균형이 맞는, 크기가 더 큰 데이터셋을 사용할 수 있었다면 성능이 더 좋아졌을 것 같습니다. 중요도1인 공지는 공지일만 추출하면 되어서 잘못 추출하는 일이 거의 발생하지 않았습니다. 그러나 중요도2인 경우 공지일과 (있는 경우) 마감일까지 추출해야 해서 정확하지 않고 약 50% 확률로 정답을 맞췄습니다. 학습 과정에 공지일과 마감일은 사용하지 않았고 api를 사용해서 추출한 것이 원인인 것 같습니다. 만일 중요도2인 공지를 대상으로 공지일과 마감일까지 학습시켰다면 실제 사용에서 더 성능이 좋을 수 있었겠지만 중요도2인 데이터셋의 개수도 적고 이 중 마감일까지 있는 데이터셋은 더 적어서 학습을 시켰어도 오버피팅이 발생했을 것 같습니다.

직접 데이터 수집, 라벨링, 전처리부터 진행해서 의미 있는 프로젝트였습니다. 데이터 전처리 중 반복되는 공지사항 제거를 중복되는 단어 개수나 코사인 유사도로 자동화하려 했으나 중요한 데이터까지 지워져 수동으로 하다 보니 라벨링과 데이터 전처리에만 상당한 시간이 소요되었습니다. 데이터셋의 크기가 작은 관계로 하이퍼파라미터 조정을 여러 번 했는데 이 과정에서 기존에 사용하던 learning rate등 말고 새로운 하이퍼파라미터들에 대해 알아볼 수 있는 기회가 되었습니다.