**K-디지털 AI 팀 프로젝트 계획서**

**수행기관: 휴먼교육센터 천안**

**과정명: 인공지능(AI) 서비스기반 웹개발자 심화 프로젝트 과정**

**1. 기본 정보**

• 팀명: 남희 신

• 팀원: 최영범(팀장), 박현재, 이원희, 이성하, 엄정민

• 프로젝트 기간: 2025년 9월 10일(수) ~ 2025년 9월 24일(수)

**2. 프로젝트 개요**

• 프로젝트 제목: 한국어 스팸/정상 메일 자동 분류 모델 개발

• 배경 및 필요성:  
 이메일은 업무와 개인 커뮤니케이션의 핵심 수단이지만, 광고·피싱·악성 링크가 포함된 스팸 메일이 현재 지속적으로 유입되고 있으며 앞으로도 계속 유입될 가능성이 높습니다. 이러한 스팸 메일은 보안 위험과 업무 효율 저하를 초래하며, 사용자의 신뢰를 떨어뜨리는 요인이 됩니다. 따라서 한국어 데이터에 최적화된 경량형 스팸 메일 자동 분류 모델을 개발하여, 누구나 쉽게 사용할 수 있는 API 형태로 제공하는 것이 본 프로젝트의 목적입니다.

**3. 목표 및 기대 효과**

• 프로젝트 목표:  
- 한국어 이메일 제목+본문을 입력받아 스팸(1)/정상(0) 자동 분류  
- 2주 안에 데이터 전처리 → 모델 학습 → API 서비스화 → 발표/데모까지 완료

• 기대 효과:  
- 초보 개발자 팀이 AI 프로젝트의 전 과정을 경험할 수 있음  
- 현업 환경에서도 재사용 가능한 경량 스팸 필터링 모듈 확보  
- AI 프로젝트 수행 경험을 기반으로 향후 딥러닝/고도화 프로젝트 확장 발판 마련

**4. 데이터 정보**

• 데이터 출처:  
- KoInFoBench 한국어 스팸 메일 데이터셋  
- AIHub 국가기록물 대상 초거대 AI 학습용 말뭉치 데이터

• 데이터 설명:  
- 이메일 제목(subject), 본문(body), 라벨(label: 1=스팸, 0=정상)

• 데이터 전처리 계획:  
- 중복 제거(제목+본문 해시)  
- 불필요한 URL/특수문자 제거  
- 학습/검증/테스트 8:1:1 분리(Stratified Split)  
- 불균형 데이터 → class\_weight="balanced" 적용

**5. 사용 기술 및 도구**

• 개발 환경: Python 3.13.7, VSCode/Jupyter, Github

• 프로그래밍 언어 및 라이브러리:  
- pandas, scikit-learn, joblib, matplotlib  
- FastAPI, Uvicorn (API 서버)

• 모델 기법:  
- TF-IDF 기반 피처 추출  
- 로지스틱 회귀(Logistic Regression)  
- 성능 비교: Naive Bayes, LinearSVC, SGDClassifier

**6. 프로젝트 수행 계획 (WBS)**

|  |  |  |
| --- | --- | --- |
| 일정 | 주요 작업 내용 | 담당자 |
| 9/10 | 주제 선정 | PM/QA |
| 9/11~14 | 데이터 수집 및 전처리 (중복 제거, 클린징, 분리) | 이원희 |
| 9/15~16 | 베이스라인 모델 학습 (TF-IDF+LogReg) 및 성능 측정 | 전원 |
| 9/17 | 오류 분석(오탐/미탐 사례 정리) | 최영범 |
| 9/18~19 | 규칙 기반 특징 추가, 모델 성능 개선, 교차검증 | 엄정민 |
| 9/22 | FastAPI 서버 구축 (/predict), 간이 데모 UI 제작 | 이성하 |
| 통합 테스트 (엣지 케이스 포함), 임계값 조정 | 전원 |
| 9/23 | 최종 보고서/모델카드 작성, 발표자료 초안 | 박현재 |
| 최종 리허설 (발표 및 데모 점검) | 전원 |
| 9/24 | 최종 발표 | 최영범 |

**7. 성능 평가 지표 및 기준**

• 사용 지표: Accuracy, Precision, Recall, F1-score, ROC-AUC

• 기준 목표:  
- F1 ≥ 0.92  
- ROC-AUC ≥ 0.97  
- 정상 메일 오탐률 ≤ 2.5% 또는 스팸 미탐률 ≤ 5%

**8. 예상 산출물**

• 최종 모델 및 코드: 학습 스크립트, 추론 스크립트, API 서버 코드

• 발표자료 (PPT): 10~12장, 프로젝트 개요·데이터·모델링·성과·시연 포함

• 기타: 모델카드(데이터 출처, 성능, 한계/바이어스, 재현 방법), 데모 UI