

2차 단위 프로젝트 개요

기간

- 기간: 11월 25일 ~ 11월 26일
- 발표: 11월 26일 4시 ~
 - 팀 당 10분 내외.
- 산출물 제출: 11월 26일 까지

프로젝트 주제: 가입 고객 이탈 예측

- 다양한 서비스에 가입한 고객이 향후 일정 기간 내에 이탈할 가능성을 사전에 예측하는 모델을 개발하는 것을 목표한다.

프로젝트 목표:

- 비즈니스 문제를 명확히 이해하고 이를 효과적으로 예측하는 머신 러닝 모델을 개발한다.

세부 수행 내용

- 프로젝트 주제와 관련된 비즈니스 목표 설정 및 이해
- 머신 러닝 모델 구축을 위한 데이터 셋 준비 및 전 처리
 - 수집한 데이터에 대한 탐색적 분석
 - 데이터 전 처리
- 학습용 데이터 셋과 테스트 데이터 셋 분리
- 머신 러닝 모델과 딥러닝 모델 학습 및 평가
 - 활용할 머신 러닝/딥러닝 모델 후보 선정
 - 모델 선정을 위한 평가 지표 선정
 - 모델링 진행
 - 각 모델들의 성능 평가 측정 지표를 바탕으로 최적의 모델 선정
- 최적의 모델 최종 테스트 및 배포

산출물

- github repository에 **산출물** 폴더를 만들고 거기에 다음 **두 개 산출물 문서** 저장
 - 데이터 전 처리 결과서
 - 모델 학습 결과서

데이터 전 처리 결과서

- 수집한 데이터에 대한 설명
 - 수집 목적
 - 수집한 방법
 - Feature들에 대한 설명
- 학습 데이터에 대한 **탐색적 데이터 분석(Exploratory Data Analysis, EDA)** 수행 결과
- 결측치 처리 방법 및 이유
- 이상치 판정 기준과 처리 방법 및 이유
- 적용한 Feature Engineering 방식

모델 학습 결과서

- 최종 모델 선정을 위한 평가 지표와 그에 대한 설명
- 최종 선정 된 모델에 대한 설명
- 학습 과정 기록
 - 어떤 모델들을 학습했고 그 성능은 어느 정도인지에 대한 설명
 - 하이퍼파라미터 튜닝 과정에 대한 설명
 - 최종 모델과 최종 평가 지표에 대해 기술

학습된 모델과 Service application

- 최종 모델을 저장한다.
- 최종 모델을 이용해 추론하는 application을 streamlit을 이용해 구현한 코드.

Github Repository

- <https://github.com/orgs/SKNETWORKS-FAMILY-AICAMP/repositories>
 - 여기에 프로젝트 Repository 생성
 - 이전 과정에서 진행한 프로젝트 참고할 수 있다.
- Repository 이름
 - **SKN기수-2nd-NTeam**
 - N: 팀 번호
 - 예: SKN21-2nd-1Team
- GIT Repository
 1. 팀장이 repository를 생성한다.
 2. 팀원을 협력자(collaborator)로 등록한다.
 1. Repository의 **Settings > Collaborator and teams** 메뉴 선택
 2. Manage access의 add people 버튼 클릭.
 3. 팀원 id로 찾아서 등록

Project README.md

- Readme.md는 발표 자료로 사용할 수 있도록 작성한다.
1. 팀원 및 담당 업무
 2. 프로젝트 주제 및 주제 설명 및 선정 이유
 1. 주제
 2. 주제를 선택한 이유
 3. 주요 기능
 3. 프로젝트를 구성하는 디렉토리들과 파일들의 구조.
 4. 수집 데이터 설명
 5. 데이터 베이스 테이블 설명
 6. Application의 주요 기능
 - Application 제공하는 기능 설명
 7. 회고
 1. 구현 도중 생겼던 문제와 어떻게 해결했는지 작성.
 2. 각자 프로젝트 진행하면서 느낀 점