

팀프로젝트 개요

2025-1 기계학습

이현석 교수

조교: 김규식, 박성호

Department of Intelligent Mechatronics Engineering,
Sejong University

- 팀프로젝트 설명
- Datasets
 - Description 예시
- 평가 방법

■ 산업 현장에서 활용할 수 있는 기계학습 기반 데이터 분석 및 모델 개발

- 세부 주제의 경우 팀별로 직접 선정 (주어진 데이터셋 이용한 주제 또는 직접 찾은 자유 주제)

■ 목표

- 기계학습을 활용한 문제 해결 팀프로젝트 수행
- 딥러닝을 제외한 기계학습 기법만을 활용해서 문제 해결
- 팀원과의 협업을 통해 데이터 분석 및 모델 개발 과정 경험
- 다양한 데이터셋을 활용한 창의적인 주제 선정 및 맞춤형 기계학습 모델 설계
- 산업 현장에서 적용 가능한 기계학습 모델의 한계점 및 개선 방향 분석

■ 팀 구성

- Random 3인 1팀
 - 총 수강 인원 42명, 14팀

■ 과제 및 프로젝트(40%)

- 과제 (10%)
- 프로젝트 (30%) - 중간평가(5%), 최종평가(25%)

■ 중간평가(5%)

- 11 or 12주차 강의시간에 5분 내외 팀별 발표 (일정은 강의 상황 고려하여 조정)
 - 팀별 세부 주제와 Dataset의 적합성, 프로젝트 방향성 등 평가
- 발표 자료 제출

■ 최종평가(25%)

- 15주차 강의시간에 10분 내외 팀별 발표
 - 데이터 분석 수준, 문제 해결 접근 방법의 적절성, 한계점 및 개선 방향 등 평가
- 최종 발표 자료 및 개인별 보고서 제출, 팀원 평가
 - (예정) 최종 발표(15%), 개인별 보고서(5%), 팀원 평가(5%)
 - 개인별 보고서는 팀 주제 및 문제 해결 내용 + 본인이 어떤 기여를 했는지 작성

■ 기본적으로 제공하는 5가지 Datasets

- 자유 주제의 경우 다른 데이터셋을 찾아서 사용

■ Datasets 및 주제 예시

- Advertising dataset
 - 여러가지 광고 비용에 따른 총 매출 예측
- Bank marketing dataset
 - 고객에 대한 다양한 특징을 기반으로 은행 정기 예금 가입 여부 예측
- Car evaluation dataset
 - 차량의 다양한 특징을 기반으로 평가 수준을 예측
- Diabetes dataset
 - 당뇨병 환자에 대한 여러 특징을 기반으로 환자의 1년 후 당뇨병 진행 정도 예측
- Medical cost personal dataset
 - 피보험자의 여러 특징을 기반으로 개인 의료비 청구 금액 예측

■ 각 Dataset 마다 Description 참고

– Datasets – (ex)Bank_Marketing

| 이름 | 수정된 날짜 | 유형 | 크기 |
|---------------------------------|-------------|--------------------|---------|
| Bank_Marketing_columns_Info.csv | 2025-03-... | 한컴오피스 한셀 ... | 1KB |
| Bank_Marketing_Dataset.csv | 2025-03-... | 한컴오피스 한셀 ... | 5,068KB |
| Bank_Marketing_Dataset_Info.pdf | 2025-03-... | Adobe Acrobat D... | 272KB |

데이터셋 특성

| 특성 | 설명 |
|--------|---------------------|
| 다변량 | 예 |
| 주제 분야 | 비즈니스 |
| 연관된 작업 | 분류 (Classification) |
| 특성 유형 | 범주형, 정수형 |
| 인스턴스 수 | |
| 특성 수 | |

데이터셋 정보

추가 정보

이 데이터는 포르투갈의 한 은행에서 진행한 직접 다. 종종 같은 고객에게 여러 번 연락을 해야 했을지 ('no') 예측하는 것이었습니다.

데이터셋은 총 네 가지로 구성되어 있습니다:

1. **bank-additional-full.csv**: 전체 예제(41, 년 5월에서 2010년 11월까지). [Moro et al
2. **bank-additional.csv**: 1)에서 무작위로 선택
3. **bank-full.csv**: 전체 예제와 17개의 입력 변
4. **bank.csv**: 3)에서 무작위로 선택된 10%의

| 변수명 | 역할 | 유형 | 인구통계학적 정보 | 설명 | 단위 | 결측치 |
|-------------|----|-----|-----------|--|----|-----|
| | | | | 'illiterate', 'professional.course', 'university.degree', 'unknown') | | |
| default | 특성 | 이진형 | | 신용 불량 여부 | | 없음 |
| balance | 특성 | 정수형 | | 연간 평균 잔고 | 유로 | 없음 |
| housing | 특성 | 이진형 | | 주택 대출 여부 | | 없음 |
| loan | 특성 | 이진형 | | 개인 대출 여부 | | 없음 |
| contact | 특성 | 범주형 | | 연락 방법 (범주형: 'cellular', 'telephone') | | 있음 |
| day_of_week | 특성 | 날짜형 | | 마지막 연락 요일 | | 없음 |
| month | 특성 | 날짜형 | | 마지막 연락 월 (범주형: 'jan', 'feb', 'mar', ..., 'nov', 'dec') | | 없음 |
| duration | 특성 | 정수형 | | 마지막 연락 지속 시간 (초 단위) | | 없음 |
| campaign | 특성 | 정수형 | | 이번 캠페인에서 고객에게 이루어진 연락 횟수 (마지막 연락 포함) | | 없음 |
| pdays | 특성 | 정수형 | | 이전 캠페인에서 마지막 연락 이후 지난 일수 (-1은 이전에 연락을 받은 적 없음) | | 있음 |
| previous | 특성 | 정수형 | | 이전 캠페인에서 고객에게 이루어진 연락 횟수 | | 없음 |

감사합니다
