

스프린트 미션 15

미션 개요

- 두 명의 연구자가 협업하는 도커 기반 워크플로우를 설계
 - 연구자 1 (Trainer): 데이터 전처리, 모델 학습, 모델 파일 추출
 - 연구자 2 (Lab): 학습된 모델을 활용한 추론 및 분석
-

데이터 처리 및 모델링

- 전처리
 - 중복 제거: train 64개, test 75개
 - 범주형 처리: *Extracurricular Activities* → Yes=1, No=0 (Int8 인코딩)
- 모델링
 - Input Features: *Hours Studied, Previous Scores, Extracurricular Activities, Sleep Hours, Sample Question Papers Practiced*
 - Target Variable: *Performance Index*
 - 파이프라인: *StandardScaler + Ridge Regression*
 - 하이퍼파라미터 탐색: $\alpha \in \{0.01, 0.1, 1.0, 3.0, 10.0, 30.0, 100.0\}$
- 성능
 - Best α : 0.1
 - CV Best RMSE: 1.4312
 - Holdout RMSE: 2.0375
- Output
 - 모델: `model.pkl`
 - 메트릭: `metrics.json`

Docker 워크플로우

- Trainer 컨테이너
 - 연구자 10이 만든 Docker 이미지 기반 실행
 - `train.py` 수행 후 `model.pkl`, `metrics.json`, `mission15_test.csv` 를 `./shared` 폴더에 저장
- Lab 컨테이너
 - 동일 이미지를 사용하여 Jupyter Lab 실행
 - `./shared` 폴더에서 trainer의 결과물을 활용
 - `./notebooks` 폴더를 마운트하여 로컬 노트북 파일 관리 가능
 - 포트 8888 연결 후 브라우저에서 Jupyter Lab 접속
 - Jupyter Lab 환경에서 `inference.ipynb` 를 작성·실행 → `result.csv` 파일 생성

Docker Hub

- 이미지 빌드 & 푸시

```
docker build -t daeseok2233/codeit_mission15:latest .
docker push daeseok2233/codeit_mission15:latest
```

- Docker Hub URL

https://hub.docker.com/repository/docker/daeseok2233/codeit_mission15