
이수민

Applied Data Scientist(기술경영 및 산업 데이터 분석 기반)

✉ milpasoomin@gmail.com

🌐 github.com/Leesoomin97 | 📄 https://blog.naver.com/milpa

📍 Based in Seoul | Remote 가능

자기소개

안녕하세요.

식품공학과 기술경영을 다전공하며 기술과 산업 데이터를 경영적 관점에서 분석해온 데이터 사이언티스트 이수민입니다.

대학교 졸업 후 감정평가사를 준비하며 데이터 기반 의사결정과 가치 분석의 중요성을 실감했고, 이후 AI 산업의 성장 가능성을 보고 AI 부트캠프에 참여했습니다.

Python, SQL, Docker, Airflow, FastAPI 등 다양한 환경에서 데이터 수집·모델링·자동화 전 과정을 실습하며 실무 역량을 쌓았고,

이를 바탕으로 Dacon Toss CTR 예측 대회와 MLOps 기반 게임 추천 프로젝트 등을 수행하며 데이터 기반 문제 정의부터 모델링·검증·적용까지의 실무 과정을 경험했습니다.

앞으로도 빠른 학습력과 실행력을 기반으로 폭넓은 지식을 실무에 적용해, 조직의 의사결정 효율을 높이는 데이터 사이언티스트로 성장하고자 합니다.

핵심 역량 및 스킬

1. 데이터 분석 및 모델링 역량

- ❖ CTR 예측 정확도 향상을 위해 **10M+** 로그데이터 기반 모델링 파이프라인 설계 및 최적화
 - ❖ 다양한 데이터 패턴 비교를 위해 **LightGBM, XGBoost, DCN** 구조 병행 및 앙상블 전략 적용
 - ❖ 클릭 행동의 맥락 반영을 위해 세션 길이, 다양성 비율, 시간·지면 교차 피처 등을 직접 설계
 - ❖ 지역·시점별 가격 변동 분석을 위해 시계열·공간 변수를 결합한 아파트 실거래가 회귀 모델 개발
-

2. 데이터 서비스 운영 및 기술 인프라 이해 역량

- ❖ 데이터 수집-학습-배포 전 과정 자동화를 위해 **Airflow** 기반 워크플로우 파이프라인 도입
 - ❖ 모델 예측 결과를 실시간으로 제공하기 위해 **FastAPI·Docker** 기반 서빙 및 웹 프로토타입 구현
 - ❖ 운영 안정성을 위해 **AWS S3·MySQL** 기반 데이터·로그 관리 체계 구축 및 배포 환경 최적화
 - ❖ 효율적 협업을 위해 **Git·GitHub Actions** 기반 CI/CD 자동화 환경 구성 및 코드 검증 절차 개선
-

3. 문제 해결 및 성장 역량

- ❖ 빠른 기술 적응을 위해 **Airflow, Docker, GitHub Actions** 등 신규 도구 자율 학습 후 실무 적용
- ❖ 데이터 품질 및 흐름 개선을 위해 문제 원인 분석 후 구조적 개선안 제시
- ❖ 프로젝트 완성도 제고를 위해 새로운 환경에서도 빠르게 학습·적용·성과 도출

프로젝트 요약

[ML] Toss NEXT ML Challenge : 광고 클릭 예측(CTR) 모델 개발
10M+ 규모 광고 로그데이터를 기반으로 LightGBM·XGBoost·DCN 모델을 병행한 CTR 예측 모델 개발.
세션·시간·노출 맥락을 반영한 피처를 직접 설계하고, k-fold 활용 및 Weighted 앙상블(45:55)로 예측 안정성 향상.
(Python, Scikit-learn, LightGBM, XGBoost, PyTorch)

[MLOps] Game Recommendation System : RAWG API 기반 게임 추천 서비스
RAWG API로 수집한 게임 데이터를 활용해 Item-based CF 추천 모델 구축.
Airflow·FastAPI·Docker·AWS S3 기반 파이프라인을 설계해 데이터 수집-학습-배포 전 과정 자동화.
(Python, LightFM, Airflow, FastAPI, MySQL, Docker, GitHub Actions)

교육 / 자격 / 활동

1. 교육

- ❖ 건국대학교(서울) — 축산식품생명공학과 전공 / 기술경영학 다전공
 - › 재학기간: 2016.03 - 2021.02
 - › 졸업 구분: 4년제 학사 졸업
 - › 주요 이수 과목: 기술경영론, 기업기술가치평가론, 전략경영론, R&D관리, 지식재산 애널리틱스 등
 - › 기술·경영 융합형 인재로 성장하기 위해 공학적 데이터 이해와 경영 의사결정 과정을 병행 학습
- ❖ **AI Bootcamp 16기 (Upstage)**
 - › 기간: 2025.07 - 2025.11 (진행 중)
 - › 주요 내용: MLOps 실습, Airflow & FastAPI 기반 추천 시스템 구축, 대용량 데이터 분석, ML/DL 모델링 실습
 - › Python, SQL, Docker, AWS S3, GitHub Actions 등 실무 환경에서 데이터 수집·모델링·배포 전 과정을 경험

2. 자격 / 어학

- ❖ 감정평가사 1차 2회 합격 (2022, 2024)
 - › 경제학, 회계학, 민법, 부동산학 기반의 가치평가 이론 학습을 통해 데이터 기반 의사결정 및 정량 분석력 강화
- ❖ **TOEIC 755점 (2024.01)**
- ❖ 2종보통운전면허(오토) — 경찰청(운전면허시험관리단), 2018.09

❖ (준비 중)정보처리기사 필기 합격, 빅데이터분석기사 필기 합격, ADsP (데이터분석 준전문가) 준비 중, SQLD (SQL 개발자) 준비 중

3. 대외활동

- ❖ 한국투명성기구 청년기자단 (2018.04 - 2018.12)
 - › 사회 이슈 조사 및 이에 대한 데이터 기반으로 기사 작성
 - › 청년 투표 참여율, 숙명여고 사태 등 주요 주제 분석 기사 발행 및 연간 활동집 수록
- ❖ 서울시설공단 시민모니터링단 (2018.04 - 2018.12)
 - › 공공시설 이용자 불편사항 조사 및 개선안 제안
 - › 어린이대공원 영장류 철제 우리 교체 제안 → 실제 구조물 개선 반영
- ❖ 서울시장 선거대책본부 청년특보 (2018.06 - 2018.09)
 - › 청년 정책 홍보 및 SNS 콘텐츠 제작 담당
 - › 대학생 대상 선거 인식 조사 및 참여 캠페인 진행
- ❖ 차세대리더포럼 보건의료 합봉사단 (2018.02 - 2018.06)
 - › 차세대리더포럼 주최·국회 사회공헌포럼 후원 'HYPHO 보건의료 합봉사단' 활동
 - › 해방촌 지역 어르신 대상 가정 방문 건강 모니터링 프로젝트 수행
 - › 혈압·당뇨 확인, 주거 환경 및 복지 서비스 필요 여부 조사 등 데이터 기반 점검 진행
 - › 원활한 커뮤니케이션 능력을 인정받아 차기 봉사단원 대상 활동 절차 및 대화 방법 강연 진행
 - › 데이터 활용 역량과 리더십·소통 능력을 동시에 강화한 경험
- ❖ 기타 활동
 - › 건국대학교 PRIME 사업단 창업 *K-Food Lab* 인턴십 — 비건 만두 개발 아이템 사업 기획 및 수익성 분석 (2017.09-2017.12)
 - › 건국대학교 합창단 — 신입생 공연 책임자 및 알토 파트 단원으로 정기공연 참여 (2017.03-2017.12)

프로젝트 상세

[ML] Toss NEXT ML Challenge : 광고 클릭 예측(CTR) 모델 개발

- ❖ 프로젝트 형태: 개인 프로젝트
- ❖ 기간 / 인원: 2025.09.28 - 2025.10.12 (2주) / 1인
- ❖ 개요:
 - › 토스 앱 광고 로그(10M+ 행)를 기반으로 사용자의 광고 클릭 확률을 예측하는 머신러닝 모델 개발
 - › 트리 기반 모델(LightGBM, XGBoost)과 신경망(DCN)을 병행하여 구조적 특징 비교 및 앙상블 적용
 - › CTR 예측에 특화된 피처(seq_length, diversity_ratio, inventory_id_hour_cross 등) 설계로 성능 향상
- ❖ 스킬:

- › Python · Pandas · Scikit-learn · LightGBM · XGBoost · PyTorch
 - ❖ 주요 역할:
 - › 데이터 전처리 및 피처 엔지니어링 전 과정 수행
 - › 트리 기반 / 딥러닝 기반 모델 병행 학습 및 K-Fold 검증
 - › Weighted 앙상블(45:55) 적용 및 feature importance 기반 조합 최적화
 - ❖ GitHub: [Leesoomin97/toss_ctr_dacon_project](https://github.com/Leesoomin97/toss_ctr_dacon_project)
-

[MLOps] Game Recommendation System : RAWG API 기반 게임 추천 서비스

- ❖ 프로젝트 형태: 팀 프로젝트 (Team Third Party)
- ❖ 기간 / 인원: 2025.09 - 2025.10 (1개월) / 5인
- ❖ 개요:
 - › RAWG API로 게임 메타데이터를 수집해 Item-based Collaborative Filtering 추천 시스템 구축
 - › Airflow, FastAPI, Docker, GitHub Actions, MySQL, AWS S3를 연계해 데이터 수집-학습-배포 자동화
- ❖ 스킬:
 - › Python · Pandas · LightFM · Airflow · FastAPI · MySQL · AWS S3/EC2 · Docker · GitHub Actions
- ❖ 주요 역할:
 - › 데이터 전처리 및 ItemCF 모델 구현 (Recall@10 = 0.69)
 - › Airflow DAG 설계로 모델 재학습 및 추론 자동화
 - › FastAPI 기반 실시간 추천 웹페이지 개발 (PC·모바일 반응형 HTML)
 - › Docker·GitHub Actions 기반 CI/CD 파이프라인 및 AWS 배포 환경 구축
- ❖ GitHub: [Leesoomin97/mlops-game_recommendation_project_team3](https://github.com/Leesoomin97/mlops-game_recommendation_project_team3)