

Yoonseo Lee

Email: leeyoonseo@catholic.ac.kr

Mobile: (+82) 010-8604-5870

Portfolio: leeyoonseo.com



Summary

사용자 세션 데이터와 유저 메타정보를 활용한 추천시스템을 연구하면서, 대규모 데이터를 안정적으로 다루는 시스템 자체에 관심을 가지게 된 이윤서입니다. 학부 과정에서 결측치에 강건한 Gradient Boosting 모델과 음성 상담 데이터 기반 자동 매칭 시스템을 연구했고, 세션·유저·금융 데이터를 통합한 추천 모델을 설계해 콜드스타트 사용자에서의 성능을 개선했습니다. 이러한 경험을 바탕으로, 대규모·불완전 데이터 위에서 머신러닝 파이프라인이 효율적이고 견고하게 동작하도록 만드는 데이터 중심 ML 시스템 연구에 관심을 두고 있습니다.

CONFERENCE

결측치에 강건한
Gradient Boosting 모델

PATENT

음성 상담 데이터 기반
자동 매칭 시스템

PROJECT

대규모 세션 로그
기반 통합 추천 시스템

INTEREST

대규모 데이터 위
견고·효율적
ML 파이프라인

Education

• 가톨릭대학교

학사과정 (GPA: 4.08/4.5)

◦ 컴퓨터정보공학부

Seoul, South Korea

Mar. 2020 – Present

Research Interest

- AI for Systems
- Machine Learning & Data Mining
- Systems for AI
- Applied AI

Research Experiences

• Applied AI Lab, Catholic University of Korea

학부연구생 (지도교수: 장재연)

Seoul, South Korea

Feb. 2023 – Present

Conference

- 대한산업공학회 / 한국경영과학회 / 한국시물레이션학회 춘계공동학술대회 (2025) – MissGBM: 결측치에 강건한 Gradient Boosting 모델
공동연구자 / 이윤서*, 유준영*, 장재연†
 - 결측치를 대체하지 않고 학습 과정에서 각 피처의 예측 기여도와 학습 유용성을 동적으로 평가해 반영하는 boosting 기반 모델 제안
 - 부스팅 라운드마다 각 피처의 예측 기여도와 학습 유용성을 평가해 feature sampling 확률을 적응적으로 조정

Patent

- 돌봄 서비스 제공을 위한 기계학습 모델의 학습 데이터 생성 방법 및 이를 수행하는 전자 장치
발명자 / 장재연, 이윤서, 유준영, 김민식, 이흥주
 - 환자·보호자 상담 음성으로부터 환자 키워드를 추출하고 간병인 메타데이터와 결합해 매칭 모델 학습용 데이터셋을 자동 생성하는 방법 고안
 - 출원번호 제10-2024-0160868호

Projects

- 그래프 신경망 기반 금융상품 추천 연구
네이버파이낸셜 산학협력 Aug. 2025 – Present
 - 푸시 알림 채널 기반 금융상품 추천 시스템 연구
 - 문제 상황: 푸시 알림 발송 시 전환 가능성이 높은 유저를 선별해 conversion을 극대화할 필요성
 - 방법론: 유저 메타데이터와 page 이동 session log를 결합해 행동 패턴을 모델링하고, page hierarchy와 세션 간 상호작용 학습을 위해 GNN 기반 추천 모델 개발
 - 결론: 진행 중 (네이버파이낸셜 실서비스 데이터 기반)
- AI 기반 환자-간병인 매칭 시스템 개발
(주)코드블라썸 산학협력, 한국연구재단(NRF) 지원 Apr. 2024 – Nov. 2024
 - 음성 상담 데이터 기반 환자-간병인 자동 매칭 시스템 개발
 - 문제 상황: 환자·보호자 상담 내용을 정량화해 간병인과 매칭하는 자동화 파이프라인 부재
 - 방법론: OpenAI Whisper STT, LLM 기반 keyword extraction, RandomForest 기반 만족도 예측까지 5개 모듈 (SENSOR·STT·LABELING·FIT·MATCHING)로 구성된 파이프라인을 구현하고, 학습/예측 파이프라인을 분리해 데이터 축적 기반 모델 업데이트 구조 구축
 - 결론: 매칭 모델 학습 데이터 자동 생성 방법으로 특허 출원(제10-2024-0160868호)
- MIMIC-IV 기반 급성신손상(AKI) 발생 예측 모델 개발
2024 SNUBH Datathon (분당서울대학교병원) Jul. 2024 – Sep. 2024
 - ICU 환자의 시점별 AKI 발생 조기 예측 모델 개발
 - 문제 상황: ICU 입원 환자의 AKI 발생을 조기에 예측해 임상 의사결정을 지원할 필요성
 - 방법론: 임상 기준 기반 코호트 필터링 및 24시간 단위 Time Window 데이터셋 구축, MICE 보간·SMOTE 적용 후 XGBoost·CatBoost·LightGBM·AdaBoost·RandomForest를 F1 가중 앙상블로 결합
 - 결론: 후기 윈도우(168-192h)에서 F1 0.9186 달성, 연령·성별 그룹별 공정성 평가 수행

Awards

- 2026 캡스톤디자인공학부 장려상 Jan. 2026
 - 그래프 기반 콘텐츠 추천 시스템 개발
 - 문제 상황: Steam, IMDb 등 서로 다른 플랫폼의 콘텐츠를 하나의 그래프에서 연결해 개인 취향 기반 연관 추천을 제공할 필요성
 - 방법론: 콘텐츠·장르·콘텐츠 종류를 노드로 구성한 그래프를 구축해 플랫폼이 다른 콘텐츠도 동일 그래프 내에서 연결, LightGCN 기반 추천 모델 구현 후 모델 경량화를 위해 random walk 기반 임베딩 학습 방식도 비교 실험
 - 결론: 플랫폼·장르를 가로지르는 연관 콘텐츠 추천 파이프라인 구축
- 2025 부천시 & 가톨릭대학교 데이터 분석 경진대회 (최우수상) Aug. 2025
 - 가톨릭대 인근 음식점 대상 사용자 맞춤 추천 시스템 개발
 - 문제 상황: 보편적 평점이 아닌 사용자별 맞춤 추천을 제공할 수 있는 추천 시스템 필요성

- 방법론: 코사인 유사도 기반 user/item 하이브리드 협업필터링, FunkSVD + XGBoost 2단계 모델, GMF와 MLP를 결합한 Neural CF 세 가지 추천 모델 구현·비교
- 결론: 유저 10명 대상 시나리오 평가에서 긍정적 반응 확인, 콜드 스타트 환경에서도 효율적으로 동작함을 검증

- **2023 해군·해병대 AI 경진대회 (참모총장상, 본선 3위)**

Jul. 2023

- 드론 객체 탐지 모델 개발
- 문제 상황: 동영상 프레임 캡처 이미지에서 드론의 크기가 초소형부터 대형까지 제각각으로, 단일 모델로는 모든 스케일을 안정적으로 탐지하기 어려운 한계
- 방법론: YOLOv5 기반 객체 탐지 모델을 객체 크기별(소형·중형·대형)로 분리 학습한 뒤 앙상블로 결합하고, EDA를 통한 이미지 밝기 보정·불량 이미지 필터링·데이터 증강으로 정밀도 개선
- 결론: 약 40명의 참가자, 본선 4인 1팀 중 3위로 참모총장상 수상

Skills

- **Programming** – Python, SQL, JavaScript
- **Frameworks** – PyTorch, scikit-learn
- **Machine Learning** – Recommender Systems, Graph Neural Networks, Gradient Boosting
- **Tools** – Git, Docker, AWS

Community

- **Kakao x Goorm 구름톤유니브 – 교내 대표**

Jul. 2024 – Dec. 2024

- 전국 50개 대학 지부를 둔 IT 개발 연합 동아리의 교내 대표로 14명 규모 챗터 운영 총괄
- 4개 교내 IT 동아리 합동 해커톤(80명) 및 15개 대학 연합 판교 해커톤(100명) 기획·개최, 학사 행정과 협의해 수상팀에 졸업요건 대체 인정 부여

- **AWS Cloud Clubs (ACC)**

Jan. 2024 – May. 2024

- AWS 공식 대학생 커뮤니티 2nd 기수 수료. EC2, RDS, S3, IAM, Lightsail 등 클라우드 인프라 서비스의 동작 원리와 활용 방법 학습

- **Google Developer Student Clubs (GDSC)**

Jan. 2024 – Jul. 2024

- Google 공식 대학생 커뮤니티에서 알고리즘 스터디 활동(DP, DFS/BFS, Dijkstra, Kruskal, SCC). 운영 중인 매칭 알고리즘에 이진탐색트리를 도입해 시간복잡도를 $O(n^2)$ 에서 $O(k \log n)$ 로 개선

- **University Makeus Challenge (UMC)**

Nov. 2023 – Mar. 2024

- 25개 대학 IT 연합 동아리에서 Node.js 기반 서버 개발 학습. ERD 설계·MySQL·Swagger API 명세·CI/CD를 적용해 온라인 타임캡슐 서비스 백엔드 개발 및 런칭