



김 광 민

Data Scientist/ Data Analyst

Email: kmink3225@gmail.com

Website: kmink3225.netlify.app

Linkedin: <https://www.linkedin.com/in/kwangmin-kim-a5241b200/>

ABOUT ME

통계 및 machine learning에 대한 전문 지식을 갖춘 data science 분야에서 7년여 동안의 data analysis 경험이 있으며 R, Python, SQL 등과 같은 오픈 소스 도구를 사용합니다. 생화학, 수학 및 생물 통계학을 전공했고, modeling 및 data analysis를 통해 얻은 상세한 객관적 사실을 바탕으로 비전문가와 소통할 수 있습니다. 업무 방식은 계획적이고 체계적으로 일하는 것을 목표로 합니다.

SKILLS

Data Science

- R, Python
- SAS

Database

- SQLite
- Oracle-SQL

etc

- Ubuntu, Powershell, Git/Github, Conda
- Quarto, R markdown, Jupyter,

EXPERIENCE

2020.12 - 재직 중

(주) 씨젠, 진단 IT 총괄 연구소, Data Science Team

Data Scientist / Data Analyst

- 진단 신호 처리 알고리즘을 위한 DHF(Design History File) 기획 및 집필했습니다. 현재 해외 각국의 규제 요구 사항을 준수하기 위해 문서를 업데이트하고 관리하고 있습니다. 이 프로젝트는 유럽, 북미, 남미, 남아프리카, 서아시아를 포함한 글로벌 비즈니스에 반드시 필요한 프로젝트로 분기당 약 1,000억 원(7,500만 달러)의 매출을 기록하고 있습니다.
- 통계적 테스트를 이용한 진단 신호 처리 알고리즘의 안전성에 대한 FDA 검증 및 인허가 보고서 문서를 기획 및 집필하여 1건의 특허 발명을 달성했습니다. 현재 FDA 규제 요구 사항을 준수하기 위해 문서를 업데이트하고 관리하고 있습니다. 이 프로젝트는 미국 시장 진출을 위한 SG의 최우선순위 프로젝트입니다.
- Levenberg-Marquardt 알고리즘을 활용한 의료기기의 신호 데이터를 처리하기 위한 진단 알고리즘을 관리 및 VOC (Voice of Customer) feedback을 수행합니다. 이 진단 알고리즘으로 SG는 2.5년 동안 약 2조 6천억 원(20억 달러)의 매출을 기록했습니다.
- 실험자, 전략기획자, 임원, 변리사 등 비전문가를 대상으로 실험 설계, 데이터 분석, 통계 분석 컨설팅 제공했습니다.
- 플랫폼 전략 IP(지적 재산)기획 TF팀에 발탁되어 아이디어 26건, 발명 16건 및 특허발명 5건 달성했습니다.
- 씨젠 Full Automation을 위한 DevOps 프로젝트로, 의료 장비를 위한 품질 관리(QC) 알고리즘을 개발하여 2개의 특허 발명을 출원했고 QC 프로세스에 소요되는 시간이 153배 이상 감소하고 6억 원(\$450,000)에 달하는 QC 비용을 13배 감소시켰습니다:
 - 장비 오류, 인적 오류 및 시약 생산 라인 오류를 분류했습니다.
 - 소프트웨어 엔지니어에게 프로젝트의 타당성을 입증하기 위한 프로토타입으로 QC 플랫폼의 웹 애플리케이션을 개발했습니다.
 - 머신러닝을 활용하여 1단계 교정 데이터에 대한 2단계 QC 과정에서 노이즈 테스트 결과를 예측하는 알고리즘을 개발하여 시간이 많이 소요되는 QC 과정을 줄였습니다.

- 의료 장비 QC 과정에서 시간이 많이 소요되는 noise test가 필요하지 않음을 통계적으로 입증하여 noise test 폐지 및 QC 과정을 간소화했습니다.

2019.05 - 2020.04

**Columbia University Irving Medical Center,
Taub Institute for Research on Alzheimer's Disease and the Aging Brain**

Research Assistant

- 데이터 QC, 결측치 분석, 통계 분석, 데이터 마이닝, 기계 학습 및 경로 분석을 포함한 파일럿 데이터를 사용하여 LLFS(Long Life Family Study)를 위한 분석 파이프라인을 구축하고 제안했습니다.
- 통계, 다음과 같은 기계 학습(ML) 및 데이터 마이닝 방법을 사용하여 시각화된 임상 데이터 분석을 수행했습니다: dimension reduction for high-dimensional data, addressing highly correlated variables through variable extraction and selection using techniques such as Lasso, ridge regression, elastic net, principal component analysis, partial least squares, and sparse-partial least squares.
- 8개월 동안 연구소에서 파악하지 못한 강력한 교란자를 데이터 마이닝을 통해 발견했습니다.

2018.12 - 2019.05

**Columbia University Irving Medical Center (CUIMC),
Taub Institute for Research on Alzheimer's Disease and the Aging Brain**

Intern

- 질병 상태를 가장 잘 예측한 분류기를 평가하여 metabolomics data에 대한 최적의 분류기를 선택하기 위해 다음의 기계 학습 방법에 대한 비교 연구를 수행했습니다: lasso, ridge regression, elastic net, decision tree, random forests, ada boosting, gradient descent boosting, support vector machine (SVM), partial least square, and sparse partial least square.
- Columbia University의 Mailman School of Public Health 연례 연구 발표회에서 포스터 발표를 했고 석사과정 대학원생을 위한 연례 연구 경진대회에서 약 100명의 대학원생 중 상위 3명으로 선정되어 상금 \$1,000와 학과장상을 수상했습니다.

2014.12 - 2015.06

The City University of New York (CUNY)

Trainee Researcher

- 미분 방정식과 비선형 최소 제곱 알고리즘을 이용하여 중금속이 찻잎에 흡착되는 과정을 반영한 mechanistic model 개발했습니다.
- 선형대수와 genetic algorithms을 이용하여 social network에서 정보 확산 모형 구축에 대한 이론적 연구 수행했습니다.
- New York City College of Technology (CUNY), BMCC (CUNY), 및 Manhattan College에서 개최되는 2015 연례 회의의 기고 논문 및 포스터 세션을 위한 발표회에서 포스터 발표를 했습니다.

EDUCATION

2017.08 - 2019.05

- Columbia University in the City of New York, Biostatistics(MS), 학과장 상 수상

2015.08 - 2017.05

- Baruch College, The City University of New York, Mathematics(BA)

2006.03 - 2012.02

- 강원대학교, 생화학(BS), 수석 졸업, 전액 장학금