



Kwangmin Kim

Data Scientist/ Data Analyst

Email: kmink3225@gmail.com

Website: kmink3225.netlify.app

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

PROFILE

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 analysis을 통해 얻은 상세한 객관적 사실을 바탕으로 비전문가와 소통할 수 있습니다. 업무 방식은 계획적이고 체계적으로 일하는 것을 목표로 합니다.

EDUCATION

2017.08 - 2019.05

Columbia University in the City of New York (CU), New York City, New York
생물통계학, Master of Science, *연례 졸업 연구 대회에서 학과장상 수상*

2015.08 - 2017.05

Baruch College, The City University of New York (CUNY), New York City, New York
수학, Bachelor of Arts

2006.03 - 2012.02

강원대학교 (KNU), 춘천, 한국
생화학, Bachelor of Science, *수석 졸업*

ACHIEVEMENT

- | | |
|-----------|--|
| 2023 | 특허 출원, (제 1발명가) 알고리즘을 위한 인허가 모형 발명, Seegene (SG) |
| 2022 | 특허 출원, (제 1발명가) 반복 측정된 Ct 값에 기초한 맞춤형 치료 방법, Seegene |
| 2022 | 특허 출원, (제 2발명가) Community Group을 위한 분자 진단 생활화 시스템, Seegene |
| 2022 | 특허 출원, (제 1발명가) 의료 플랫폼을 위한 구독 시스템, Seegene |
| 2022 | 특허 출원, (제 2발명가) 분자 진단 음성 증명서, Seegene |
| 2022 | 특허 출원, (제 2발명가) 분자 진단을 위한 예측 모델, Seegene |
| 2021 | President's Award , R&D 부문 우수상: Noise Test 자동화 시스템 구축, Seegene |
| 2021 | 특허 출원, (제 1발명가) 진단 장비의 Noise Test 자동화 시스템, Seegene |
| 2021 | 특허 출원, (제 1발명가) 의료 장비의 Noise Level 측정 알고리즘, Seegene |
| 2021 | 수료증, EN62304 - Medical Device SW Life Cycle Process Training Course, SGS |
| 2020 | 수료증, HIPPA Certification, Columbia University Irving Medical Center |
| 2019 | Job Offer , the Taub Institute, Columbia University Irving Medical Center |
| 2019 | Chair's Award , Graduation Practicum Research Competition Winner, Biostatistics, CU |
| 2018 | 자격증, SAS Certified Base Programmer, SAS |
| 2015 | Stipend , \$1,000 Mathematical Kinetic Modeling, CUNY |
| 2014 | 자격증, SIT TESOL Instruction Certification, Rennert |
| 2012 | Stipend , \$5,000 Medical Convergence Capstone Design, KNU |
| 2012 | 학장상, 성적 우수 수석 졸업, KNU |
| 2010-2011 | 전액 장학금, 성적 우수 장학금, KNU |
| 2009 | 사단장 표창, 리더쉽 경연 최우수상, 군대 |
| 2009 | 중대장 표창, 사단 검열 우수상, 군대 |

EXPERIENCE

2020.12 - Present

Seegene, Diagnosis IT General Research Institute, 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.
- 씨젠의 전사적 장기 전략 중 하나인 SG Full Automation을 위한 DevOps 프로젝트로서, 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.

2019.05 - 2020.04

Columbia University Irving Medical Center (CUIMC),

Taub Institute for Research on Alzheimer's Disease and the Aging Brain

Research Assistant

- 고차원 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 고차원 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.

2012.08 - 2014.12

Rennert, English Language School in New York City

Trainee Instructor

- English as a Second Language (ESL) Program 에서 어학 연수를 했습니다.
- Test of English as Foreign Language (TOFLE) 공부했습니다.
- SIT Teaching English to Speakers of Other Languages (TESOL) 자격증 취득했습니다.
- 주 1회 자원자 학생 30명 대상으로 영어 문법, 독해, 듣기, 및 쓰기 강의를 했습니다.

2010.06 - 2012.02

Molecular Biology Lab, Kangwon National University (KNU)

Trainee Researcher

- cell culture 와 the western blot을 이용한 단백질 정량 분석을 수행했습니다.
- 상황 버섯이 알레르기 반응에 의한 림프관 형성에 미치는 영향과 그 효능을 입증하기 위한 연구를 했습니다.
- Medical Convergence Capstone Design의 반기 행사 발표회에서 포스터 발표를 했습니다.

2008.03 - 2010.02

병역 의무

작전병

- 군사 전략 교육/행정 업무를 담당했습니다.
- 행정 업무 담당 및 훈련 계획 문서를 관리했습니다.
- 다수의 탄약 및 전술 훈련 검열을 대비했습니다.
- 분대장 파견 교육에서 사단장 표창을 수상 및 조기진급을 했습니다.
- 중대장으로부터 사단 검열 우수 표창을 받았습니다.

PROJECTS

2022.07 - present

알고리즘을 위한 DHF 문서화 기획 및 작성

Seegene, Diagnosis IT General Research Institute

- SGS EN62304 및 FDA General Principles of Software Validation 문서를 기반으로 진단 신호 처리 알고리즘 및 문서화를 위한 문서 이력 파일(DHF)을 기획하고 작성했습니다.
- 진단 신호 처리 알고리즘에 대한 검증 및 인허가 작업을 항목화 시켰습니다.
- Unit test, integration level test, system level test 및 advanced algorithm test를 포함하는 structural test를 위한 진단 신호 처리 알고리즘의 모듈을 그룹화하고 수식화했습니다.
- Quarto, R 및 Python을 사용하여 동적 문서를 만듭니다.
- 참고: https://kmink3225.netlify.app/docs/projects/dsp_validation/

알고리즘을 위한 FDA Verification & Validation 문서화

Seegene, Diagnosis IT General Research Institute

- Product manage 로서 반복 측정 분석을 위한 시스템 수준의 통계적 테스트 모델을 기획하고 통계 분석 계획서를 작성하고 있습니다.
- 데이터 엔지니어 및 생물학자와 협업하고 부서에 흩어져 있는 문서를 수집 및 요약하고 평가 메트릭을 생성하여 system level test 및 advanced algorithm test인 통계 테스트 모델을 구현합니다.

- 데이터 엔지니어링, 데이터 품질 관리 및 통계 데이터 분석을 수행합니다.
- 제품 관리자로서 통계 테스트를 사용하여 진단 신호 처리 알고리즘에 대한 FDA 검증 및 검증 보고서 문서를 작성합니다.
- Quato, R 및 Python을 사용하여 동적 문서를 작성합니다.
- 참고: https://kmink3225.netlify.app/docs/projects/dsp_validation/

2021.12 - 2022.07

Platform Planning TF

Seegene, Strategy Planning Office & In-house Patent Center

- 플랫폼을 위한 전략 기획 및 IP 기획.
- 변리사에 의해 26개의 ideation 중 16개가 발명으로 채택됨.
- 16개의 발명 중 4개가 출원됨 (나머지는 진행 중).
- 기획자와 변리사들에게 Database system, 통계 및 ML 컨설팅 statistics, and ML 컨설팅.

2021.09 - present

Data-Driven 진단 알고리즘 개발

Seegene, Future Technology Research Institute & In-house Patent Center

- 진단기기의 광학적 특성과 SG 시약의 고유 기술을 반영하여 새로운 mechanistic model을 정의했습니다.
- back-fitting 알고리즘을 이용하여 GAM (Generalized Additive Model) 으로 data-driven 진단 알고리즘 (to-be) 제시했습니다.
- data-driven진단 알고리즘과 Levenberg-Marquardt 알고리즘을 이용한 rule-based sigmoid regression 알고리즘 (as-is)과 비교 분석할 예정입니다.

2021.01 - 2021.09

장비 품질 관리 플랫폼 구축

Seegene, Diagnosis IT General Research Institute

- Project owner 로서 프로젝트 전반에 걸쳐 모든 프로세스를 주도했습니다.
- noise level을 측정하는 개선된 QC 알고리즘을 개발했습니다.
- 다른 부서와 협력하여 QC 프로세스를 자동화하고 QC 프로세스를 시각화했습니다.
- 장치 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.
- QC 과정에서 의료기기로 시간이 많이 소요되는 noise test가 필요하지 않음을 통계적으로 입증하여 QC 과정을 간소화했습니다.
- Noise test 폐지 및 2개의 특허 발명을 달성하여 QC처리 시간을 연간 132배 이상 단축하고 QC 프로세스에서 연간 약 6억 원(\$450,000)의 비용을 13배 단축했습니다.
- 참고: https://kmink3225.netlify.app/docs/projects/qc_platform/

2020.12 -present

진단 알고리즘 관리 및 개선

Seegene, Diagnosis IT General Research Institute

- 진단 신호와 SG의 고유 기술의 특성을 잘 설명하는 mechanistic modeling을 구현, 관리 및 개선합니다.
- Rule based 알고리즘으로 신호의 전처리를 수행하고 Levenberg-Marquardt 알고리즘을 활용하여 regression을 수행하여 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

2018.12 - 2020.04

Long Life Family Study (LLFS) Project

Columbia University Irving Medical Center, Taub Institute

- 통계 및 기계 학습(ML) 분석을 수행하여 알츠하이머병과 유의하게 관련된 대사 프로필을 식별했습니다.
- Mummichog 도구를 사용하여 누락된 값 분석, 통계 분석, ML 분류 및 경로 분석을

- 포함하는 분석 파이프라인을 구축했습니다.
- Metabolomics 데이터를 이용하여 알츠하이머병과 노화된 뇌에 대한 최적의 기계 학습(ML) 방법에 대한 비교 연구를 수행했습니다.
- 참고: https://kmink3225.netlify.app/docs/projects/llfs/project_description

2015.01 - 2015.06

찾일을 이용한 중금속 제거 알고리즘 개발

The City University of New York, Mathematics

- 미분 방정식 및 비선형 최소 제곱 알고리즘을 사용한 흡착 동역학 모델링 연구: 흡착 프로세스의 가설화된 기계적 모델을 통해 이 모델이 찾일이 오염된 물에서 중금속 이온을 제거할 수 있는 속도를 예측하는 데 유용함을 입증했습니다.
- 참고: https://kmink3225.netlify.app/docs/projects/heavy_metal/

2011.01 - 2011.05

상황버섯이 림프관 형성에 미치는 영향

강원대학교, 분자생물학 실험실

- 상황버섯이 알레르기 반응에 의해 유발되는 림프관 형성에 미치는 영향에 대한 연구: 알레르기 반응은 부신피질호르몬제, 항히스타민제 등의 약물로 치료하고 그 외 많은 약물이 개발되었지만 그 효능은 일시적이다. 따라서, 상황버섯은 알레르기 치료에 대체적인 역할을 할 수 있는지 연구했습니다.
- 참고: https://kmink3225.netlify.app/docs/projects/phellinus_linteus/

TEACHING

2023

- **Trainer**, Statistical Analysis, Seegene

2022

- **Mentor**, An Introduction to Statistical Learning, Seegene

2021

- **Private Tutor**, Calculus 1 (undergraduate level), CU

2021

- **Private Tutor**, Calculus 2 (undergraduate level), CU

2020

- **Private Tutor**, IBT TOFLE, New York

2020

- **Private Tutor**, GRE General Test, mathematics, New York

2019

- **Teaching Assistant**, Probability theory (master level), CU

2016

- **Teaching Assistant**, Calculus 1, 2, 3 (undergraduate level), CUNY

2015

- **Teaching Assistant**, Precalculus (undergraduate level), CUNY

2015

- **Teaching Assistant**, Statistics (undergraduate level), CUNY

2014

- **Trainee Instructor**, SIT TESOL teaching, Rennert

2014

- **Private Tutor**, IBT TOFLE, New York

SKILLS

Data Science

- R (proficient), Python (intermediate), SAS (beginner), C++ (beginner)

Dashboard

- R shiny (intermediate)

Engineering

- Airflow (intermediate)

SQL

- PostgreSQL, SQLite, Oracle-SQL, My SQL (beginner)

Documentation

- Quarto (intermediate- contribution to the open-source quarto development with [quarto issue #5508](#)), R markdown (intermediate), Jupyter (intermediate)

Language

- English (proficient in verbal and written communication)

etc.

- Ubuntu, Powershell, Git/Github, Conda, MS Office