



김광민 (Kwangmin Kim)

Data Scientist/ Data Analyst

Email: kmink3225@gmail.com

Website: kmink3225.netlify.app

Portfolio: <https://kmink3225.netlify.app/docs/portfolio/>

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

인적사항

통계 및 machine learning에 대한 전문 지식을 갖춘 data science 분야에서 6년여 동안의 업무 경험이 있으며 R, Python, SQL, Apache 등과 같은 오픈 소스 도구를 사용합니다. 생화학, 수학 및 생물 통계학을 전공했고, modeling 및 data analysis을 통해 얻은 상세하고 객관적인 사실을 바탕으로 비전문가와 소통할 수 있습니다.

학력사항

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 대회 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

보유스킬

Data Science
Dashboard
Engineering
DBMS
Documentation

Language
etc.

- R (proficient), Python (intermediate), SAS (beginner)
- R shiny (intermediate)
- Apache Airflow (intermediate)
- SQL (Intermediate): PosstgreSQL, SQLite, Oracle-SQL, My SQL
- Quarto (intermediate- contribution to the open-source quarto development with [quarto issue #5508](#)), R markdown (intermediate), Jupyter (intermediate)
- Korean (native) English (proficient)
- Ubuntu, Powershell, Git/Github, Conda, MS Office

핵심강점

- 다양한 도메인 지식의 데이터 분석, 통계 분석 및 알고리즘 최적화 능력
생화학, 수학, 통계학, 머신러닝 등 여러 분야에서의 경험과 연구 성과를 통해 복잡한 문제를 해결하는 능력을 갖추고 있습니다.
- 프로젝트 협업능력, 관리 및 리더십
프로젝트의 팀원으로서 협업을 통해 원활한 소통과 업무 조율을 비롯해, Project Leader로서 기획, 일정 관리, 업무 분배, 효율적인 리소스 운영, 기술 문제 해결 및 산출물 구현까지 전반적인 관리를 담당한 경험들이 있습니다. 최근엔 53개의 데이터베이스와 16개 부서를 아우르는 데이터 거버넌스 표준화 프로젝트를 리드하고 있습니다.
- 분석 툴과 언어 활용 능력
Python, R, SQL 과 같은 데이터 분석 도구와 Airflow와 같은 워크플로우 도구를 프로젝트에 활용한 경험들이 있습니다.
- 효과적인 커뮤니케이션과 교육 능력
다양한 교육 경험을 통해 전문지식과 복잡한 분석 결과를 비전문가의 수준에 맞춰 쉽게

풀어서 전달하여 조직 발전에 기여할 수 있습니다.

● 특허 발명 및 창의적인 문제 해결 능력

다수의 특허 출원과 발명에 기여하며, 창의적이고 혁신적인 **Idea**로 회사의 지적 재산을 강화하는 데 핵심적인 역할을 수행했고 전략적인 기여를 했습니다.

경력사항

2020.12 - Present

Seegene, Diagnosis IT General Research Institute, Data Science Team

Data Scientist / Data Analyst

2024.08 - Present

- **데이터 거버넌스:** 데이터 거버넌스 도입과 서비스 통합을 위해 16개 부서와 협력하여 53개 데이터베이스의 표준화 프로젝트 실무를 리드합니다. 단계적 데이터 표준화를 실행하고, 현행 시스템 분석을 바탕으로 데이터 품질 관리 체계와 표준안을 수립합니다. 이를 통해 부서 간 데이터 일관성 30% 향상, 데이터 통합 시간 30% 단축, 전사적 데이터 활용도 25% 증가를 목표로 합니다.

2021 - Present

- **데이터 분석 컨설팅:** 실험자, 전략기획자, 임원, 변리사 등 비전문가를 대상으로 실험 설계, 데이터 분석, 통계 분석 컨설팅을 제공합니다.

2021 - 2024.07

- **진단 알고리즘 관리:** Real Time PCR 신호의 진단 알고리즘을 관리 및 최적화했습니다. 이상 신호 및 noise에 의해 발생할 수 있는 위양성 및 위음성과 같은 진단 오류의 risk를 최소화하기 위해 rule-based 알고리즘을 통계적이고 수학적인 방식의 data-driven 알고리즘으로 점진적으로 개선합니다. 알고리즘 별 신호 처리 및 fitting 결과를 통계적으로 분석하여 알고리즘을 최적화했습니다.
- **Mechanistic modeling과 Levenberg-Marquardt** 알고리즘을 활용하여 의료기기의 신호 데이터를 처리하기 위한 진단 알고리즘을 관리 및 VOC (Voice of Customer) feedback을 수행했습니다. 이 진단 알고리즘으로 SG는 2021년~2022년 기간 동안 약 2조 6천억 원(20억 달러)의 매출을 기록했습니다.

2023

- **인허가 문서 작성:** 통계 테스트를 이용하여 진단 신호 처리 알고리즘의 안전성에 대한 FDA 검증 및 인허가 보고서를 기획 및 집필하였습니다. FDA 규제 요구 사항을 준수하기 위해 문서를 업데이트하고 관리했습니다. 이 프로젝트는 미국 시장 진출을 위한 SG의 2023년도 최우선순위 프로젝트였습니다.

2022 - 2023

- **전략 기획:** 플랫폼 전략 IP(지적 재산)기획 TF팀에 발탁되어 아이디어 26건, 발명 16건 및 특허발명 5건 달성했습니다.

2021.12 - 2022.09

- **장비 QC 프로세스 개선:** 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.

이직 사유: 긴축 경영으로 인한 IT 지원 축소와 경력 발전의 한계를 느꼈습니다. 더 다양한 데이터를 다루며 커리어를 발전시키기 위해 이직을 결심하였습니다.

2019.05 - 2020.04

Columbia University Irving Medical Center (CUIMC),

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

Research Assistant

- **데이터 분석 파이프라인 구축:** 고차원 데이터 QC, 결측치 분석, 통계 분석, 데이터 마이닝, 기계 학습 및 경로 분석을 포함한 파일럿 데이터를 사용하여 LLFS(Long Life Family Study) 프로젝트를 위한 분석 파이프라인을 구축하고 제안했습니다.
- **고급 통계 분석 및 기계 학습 활용:** 통계, 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.

- 데이터 마이닝과 **EDA**를 통한 교란 요인 발견: 8개월 동안 연구소에서 파악하지 못한 강력한 교란자를 데이터 마이닝을 통해 발견했습니다.

이직 사유: 집에 도둑이 들어 신분증을 도난 당했으며, 이후 COVID-19 으로 인한 이민국의 무기한 휴업으로 인해 신분 갱신이 불가능해져 부득이하게 한국으로 귀국하였습니다.

2018.12 - 2019.05

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

- **Metabolomics** 데이터를 위한 **classifier** 성능 비교 연구 및 최적화: 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 University** 연구 경진대회 수상: 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

- **Mechasnistic modeling**: 미분 방정식과 비선형 최소 제곱 알고리즘을 이용하여 중금속이 찻잎에 흡착되는 과정을 반영한 mechasnistic model을 개발했습니다.
- New York City College of Technology (CUNY), BMCC (CUNY), 및 Manhattan College에서 개최되는 2015 연례 모임 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 외국인 자원자 학생 30명 대상으로 영어 문법, 독해, 듣기, 및 쓰기 강의를 영어로 했습니다.

2010.06 - 2012.02

Molecular Biology Lab, Kangwon National University (KNU)
Trainee Researcher

- cell culture 와 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 했고 500만원의 상금을 수상했습니다.

2008.03 - 2010.02

병역 의무
작전병

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

프로젝트

2024.08 - Present

Data Governance를 위한 **Data** 표준화

Seegene, Diagnosis IT General Research Institute

- 역할: Project Lead
- 전사적 데이터 거버넌스 및 서비스 통합 표준화 프로젝트 주도
- 범위: 16개 부서 및 53개 데이터베이스 (Phase 1)
- 접근 방식: 기존 시스템 분석을 기반으로 한 단계적 데이터 표준화 접근법 및 데이터 품질 관리 프레임워크
- 목표 산출물:
 - 부서 간 데이터 일관성 30% 개선
 - 데이터 통합 시간 30% 단축
 - 전사적 데이터 활용도 25% 증가
- 기술: 데이터 거버넌스 프레임워크, 프로젝트 관리 방법론, 데이터 표준화 기법, 데이터 모델링, aQueryTool, ETL 프로세스, Airflow, DBMS

2024.01 - 2024.07

Baseline Fitting Algorithm 최적화

Seegene, Diagnosis IT General Research Institute

- 역할: 알고리즘 개발 연구원
- 진단 정확도 향상을 위한 시그모이드 곡선 베이스라인 피팅 알고리즘 개선 및 표준화
- 주요 기여:
 - 기존 방법을 분석하고 개선된 베이스라인 피팅 알고리즘 개발
 - 데이터 증강 및 변환을 구현하고 다양한 기법 비교 (regression model, rule-based 알고리즘, 신경망)
 - 비기술 이해관계자를 위한 직관적인 시각화 생성
 - 다양한 사내 알고리즘을 단일 표준화 솔루션으로 통합
- 성과:
 - 진단 결과의 정확도 향상
 - 조직 전반에 걸친 알고리즘 구현 효율화
 - 명확한 시각적 표현을 통한 이해관계자의 이해도 향상
- 기술: 데이터 분석, 알고리즘 개발, 머신러닝, 데이터 시각화

2022.12 - 2023.11

알고리즘을 위한 **FDA Verification & Validation** 문서화

Seegene, Diagnosis IT General Research Institute

- 역할: Project Lead
- 복미 시장 진출을 촉진하고 여러 국가의 규제 준수를 보장하기 위한 FDA 규정 준수 문서화 파이프라인 개발
- 주요 기여:
 - 반복 측정 분석을 위한 시스템 수준의 통계적 테스트 모델 설계 및 구현
 - FDA 검증 및 유효성 검사 문서 개발을 위한 cross-functional 팀 주도
 - 데이터 엔지니어링, 품질 관리 및 통계 분석 수행
 - 포괄적인 통계 분석 계획 및 평가 지표 작성
 - Quarto, R, Python을 사용한 동적 문서화 프로세스 구현
- 성과:
 - FDA 검증 및 유효성 검사 문서화 완료
 - 통계 분석의 데이터 품질 및 신뢰성 향상
 - 동적 보고서 생성을 통한 문서화 프로세스 효율화
- 기술: 통계 분석, 데이터 엔지니어링, FDA 인허가, Quarto, R, Python, Airflow, 시각화

2022.04 - 2022.10

Data-Driven 진단 알고리즘 개발

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

- 역할: 알고리즘 개발자 / 연구 과학자
- 씨젠의 고유 시약 기술을 접목한 진단 장비용 데이터 기반 신호 처리 알고리즘 개발
- 주요 기여:

- 진단 장비의 광학적 특성과 씨젠 시약 기술을 반영한 새로운 메커니즘 모델 설계
- 모델 최적화를 위한 **back-fitting** 및 **Levenberg-Marquardt** 알고리즘 구현 및 비교
- 알고리즘 성능 향상을 위해 데이터 기반 접근법과 메커니즘적 이해를 통합
- 성과:
 - 진단용 데이터 기반 신호 처리 알고리즘 개발
 - 진단 결과의 정확도와 신뢰성 향상
- 기술: 신호 처리, 메커니즘 모델링, 최적화 알고리즘(**back fitting** 및 **Levenberg-Marquardt** 알고리즘), 데이터 기반 분석

2021.12 - 2022.07

Platform Planning TF

Seegene, Strategy Planning Office & In-house Patent Center

- 역할: 전략 기획자 / 기술 컨설턴트
- 진단 산업에서 씨젠의 시장 위치와 기술 리더십을 강화하기 위한 혁신적인 플랫폼 전략 개발 및 지적 재산 포트폴리오 확장
- 주요 기여:
 - 플랫폼 전략 및 지적 재산 기획 이니셔티브 참여
 - 선정된 발명에 대한 특허 청구항 작성 지원
 - 기획자와 변리사에게 데이터베이스 시스템, 통계, 머신러닝에 대한 전문가 자문 제공
- 성과:
 - 26개 아이디어 중 16개가 변리사에 의해 발명으로 채택됨
 - 4개 발명이 특허 출원되었으며 12개가 진행 중
 - 기술팀과 법률팀 간의 협력 개선 촉진
- 기술: 전략 기획, 지적 재산 개발, 특허 청구항 작성, 데이터베이스 시스템, 통계, 머신러닝

2021.01 - 2021.09

장비 품질 관리 플랫폼 구축

Seegene, Diagnosis IT General Research Institute

- 역할: Project Lead / 알고리즘 개발자
- 의료기기의 품질 관리 프로세스를 간소화하고 자동화하여 효율성을 개선하고 비용을 절감
- 주요 기여:
 - 프로젝트 오너로서 기획부터 구현까지 전체 프로젝트 수명주기 주도
 - 노이즈 레벨 측정을 위한 향상된 **QC** 알고리즘 개발
 - **R shiny**를 사용하여 소프트웨어 엔지니어들에게 자동 **QC** 플랫폼의 실현 가능성을 보여주는 프로토타입 웹 애플리케이션 제작
 - **QC** 프로세스 시간을 크게 단축하는 노이즈 테스트 결과 예측용 머신러닝 알고리즘 구현
 - 시간이 많이 소요되는 노이즈 테스트의 중복성을 통계적으로 입증하여 프로세스 단순화 달성
 - **QC** 프로세스 자동화 및 시각화를 위한 부서 간 협업 수행
 - 다양한 오류 유형 분류: 장비 고장, 인적 오류, 시약 생산라인 오류
 - 노이즈 레벨 측정을 위한 **QC** 알고리즘 개선
 - 웹 애플리케이션 인터페이스를 통한 **QC** 플랫폼 자동화
 - 노이즈 테스트 결과를 위한 머신러닝 기반 예측 알고리즘
- 성과:
 - 2건의 특허 발명 확보
 - 연간 처리 시간 132배 이상 단축
 - 연간 비용 약 13배 감소 (6억원 / 45만 달러)
 - **QC** 프로세스에서 별도의 노이즈 테스트 필요성 제거
- 기술: 프로젝트 관리, 알고리즘 개발, 머신러닝, 통계 분석, 웹 애플리케이션 개발, 프로세스 최적화, 부서 간 협업

2020.12 -2024.06

진단 알고리즘 관리 및 개선

Seegene, Diagnosis IT General Research Institute

- 역할: 알고리즘 개발자 / 진단 알고리즘 관리자
- 목적:
 - 씨젠의 고유 기술과 의료기기 특성을 반영하는 진단 알고리즘 개발 및 관리
 - 고객 피드백을 반영한 지속적인 알고리즘 개선
- 주요 기여:
 - 진단 신호와 씨젠 고유 기술의 특성을 반영하는 메커니스틱 모델링 설계 및 구현
 - Rule-based 알고리즘을 통한 신호 전처리 프로세스 구축
 - Levenberg-Marquardt 알고리즘 기반 회귀 분석을 통한 신호 처리 최적화
 - VOC(Voice of Customer) 피드백 수집 및 분석을 통한 알고리즘 개선
 - 진단 알고리즘의 성능 모니터링 및 지속적 개선
- 성과:
 - 알고리즘 적용 제품으로 2.5년간 약 2조 6천억 원(20억 달러)의 매출 달성
 - 고객 피드백 기반의 알고리즘 개선으로 제품 신뢰성 향상
- 기술: 메커니스틱 모델링, 신호 처리, Rule-based 알고리즘, Levenberg-Marquardt 최적화, 회귀 분석, 데이터 분석, VOC 데이터 분석

2018.12 - 2020.04

Long Life Family Study (LLFS) Project

Columbia University Irving Medical Center, Taub Institute

- 역할: 연구 통계학자/ 데이터 사이언티스트
- 고급 통계 및 머신러닝 기법을 사용하여 알츠하이머병과 노화 뇌와 관련된 대사체 프로파일 연구
- 주요 기여:
 - 대사체학 데이터 분석을 위한 포괄적인 분석 파이프라인 개발 및 구현
 - 알츠하이머병 관련 대사체 프로파일을 식별하기 위한 통계 및 머신러닝 분석 수행
 - 알츠하이머병과 노화 뇌 연구를 위한 최적 머신러닝 방법의 비교 연구 수행
- 성과:
 - 알츠하이머병과 관련된 중요한 대사체 프로파일 식별
 - 강력한 교란 변수를 발견하여 6개월간의 연구 병목 현상 해결
 - 알츠하이머 연구에서 다양한 머신러닝 방법의 효과성 평가 및 비교
 - 약 100명의 경쟁자들 중 실습 연구 경진대회 우승자로 선정
 - 학과장 표창 수상(1,000달러 연구 장학금 포함), 이후 컬럼비아 대학교 타우브 연구소(알츠하이머병 및 노화 뇌 연구소)의 채용 제안으로 이어짐
- 기술: 고급 통계 분석, 머신러닝, 생물정보학, 대사체학 데이터 분석, 연구 방법론, 데이터 분석 파이프라인 개발, 결측치 분석, Mummichog 도구를 사용한 경로 분석

2015.01 - 2015.06

찾았을 이용한 중금속 제거 알고리즘 개발

The City University of New York, Mathematics

- 역할: 연구 보조원 / 수학적 모델러
 - 차 앞을 흡착제로 사용하여 오염된 물에서 중금속 이온 제거 속도를 예측하는 수학적 모델 개발
- 주요 기여:
 - 미분방정식을 사용한 메커니스틱 모델링 공식화
 - 매개변수 추정을 위한 비선형 최소제곱(NLS) 알고리즘 구현
 - 모델 예측을 검증하기 위한 실험 수행
- 성과:
 - 2015년 연례 회의의 논문 발표 및 포스터 세션에서 연구 발표 (Manhattan College, New York City College of Technology (CUNY), BMCC (CUNY))
 - 1,000달러의 연구 장학금 수상
- 기술: 수학적 모델링, 미분방정식 모델링, NLS 최적화 알고리즘, 흡착 동역학 분석,

2011.01 - 2011.05

상황버섯이 림프관 형성에 미치는 영향 분석

강원대학교, 분자생물학 실험실

●역할: 연구 보조원

- 기존 치료법의 한계를 해결하기 위한 알레르기의 대체 치료제로서 상황버섯(*Phellinus linteus*)의 가능성 탐구

● 주요 기여:

- 림프관 형성에 대한 상황버섯의 영향을 연구하기 위한 실험 수행
- 코르티코스테로이드와 항히스타민제와 같은 전통적인 치료법과 상황버섯의 효능 비교

● 성과:

- Medical Convergence Capstone Design 반기 행사에서 연구 발표
- 5,000달러의 연구 장학금 수상

● 기술: 세포 배양, 샘플 준비, 단백질 분리, 웨스턴 블롯, 검출

주요 성과

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

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
2014

- **Trainee Instructor**, SIT TESOL teaching, Rennert
- **Private Tutor**, IBT TOFLE, New York