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/

Profile

Experienced Data Scientist with 6 years of expertise in statistical analysis and machine learning. Proficient in open-source tools including R, Python, SQL, and Apache frameworks. Academic background in biochemistry, mathematics, and biostatistics. Demonstrated ability to effectively communicate complex analytical findings and model results to non-technical audiences, translating data-driven insights into actionable recommendations.

Education

2017.08 - 2019.05 **Columbia University in the City of New York (CU),** New York City, New York
Biostatistics, Master of Science, *the Chair's merit in the Annual Research Competition*

2015.08 - 2017.05 **Baruch College, The City University of New York (CUNY),** New York City, New York
Mathematics, Bachelor of Arts

2006.03 - 2012.02 **Kwangwon National University (KNU),** Chuncheon, South Korea
Biochemistry, Bachelor of Science, *Summa Cum Laude*

Skills

Data Science
Dashboard
Engineering
DBMS
Documentation

- 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)

Language
etc.

- Korean (native) English (proficient)
- Ubuntu, Powershell, Git/Github, Conda, MS Office

Strengths

- **Statistical Analysis, and Algorithm Optimization Across Diverse Domains**
  With experience and research achievements across biochemistry, mathematics, statistics, and machine learning, I have developed strong capabilities in solving complex problems across various fields.
- **Project Collaboration, Management, and Leadership**
  Demonstrated strong collaborative skills as a team member through effective communication and task coordination. As a Project Leader, I have comprehensive experience in project planning, schedule management, task distribution, efficient resource operation, technical problem-solving, and deliverable implementation. Currently leading a data governance standardization project encompassing 53 databases and 16

departments.

- **Proficiency in Analytical Tools and Programming Languages**
  Extensive experience utilizing data analysis tools such as Python, R, SQL, and workflow management tools like Airflow in various projects.
- **Effective Communication and Training Skills**
  Through diverse training experiences, I can effectively communicate complex technical knowledge and analytical results to non-experts in an accessible manner, contributing to organizational development.
- **Patent Innovation and Creative Problem-Solving**
  Played a key role in strengthening the company's intellectual property rights through multiple patent applications and inventions, making strategic contributions through creative and innovative ideas.

Experiences

2020.12 - Present

Seegene, Diagnosis IT General Research Institute, **Data Science Team**
Data Scientist / Data Analyst

2024.08 - Present

- **Data Governance:** Leading the practical implementation of a data standardization project involving 53 databases in collaboration with 16 departments for data governance implementation and service integration. Execute phased data standardization and establish data quality management frameworks and standards based on current system analysis. Targeting 30% improvement in cross-departmental data consistency, 30% reduction in data integration time, and 25% increase in enterprise-wide data utilization.

2021 - Present

- **Data Analysis Consulting:** Provide experimental design, data analysis, and statistical analysis consulting to non-experts including researchers, strategic planners, executives, and patent attorneys.

2021 - 2024.07

- **Diagnostic Algorithm Management:** Managed and optimized diagnostic algorithms for Real-Time PCR signals. Progressively improved rule-based algorithms to data-driven statistical and mathematical algorithms to increase interpretability and to minimize diagnostic errors such as false positives and false negatives caused by abnormal signals and noise. Optimized algorithms through statistical analysis of signal processing and fitting results for each algorithm.
- Managed diagnostic algorithms for processing medical device signal data using a mechanistic modeling and Levenberg-Marquardt algorithms, and conducted VOC feedback. With these diagnostic algorithms, SG achieved sales of approximately 2.6 trillion won ($2 billion) during 2021-2022.

2023

- **Validation Documentation:** Planned and authored FDA validation and regulatory reports on the safety of diagnostic signal processing algorithms using statistical tests. Updated and managed documentation to comply with FDA regulatory requirements. This project was SG's top priority project for 2023 for entering the U.S. market.

2022 - 2023

- **Strategic Planning:** Selected for the Platform Strategy IP (Intellectual Property) Planning Task Force, achieving 26 ideas, 16 inventions, and 5 patent inventions.

2021.12 - 2022.09

- **Equipment QC Process Improvement**: As a DevOps project for SG Full Automation, one of Seegene's enterprise-wide long-term strategies, developed quality control (QC) algorithms for medical equipment, resulting in 2 patent applications. This led to a 153-fold reduction in QC process time and a 13-fold reduction in QC costs, amounting to 600 million won ($450,000).

Reason for Job Changes: Looking to expand my expertise with diverse data environments and advance my career trajectory, following strategic changes in current organization's IT investments.

2019.05 - 2020.04

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

- **Data Analysis Pipeline Development**: Developed and proposed a data analysis pipeline for the Long Life Family Study (LLFS) project using pilot data, incorporating high-dimensional data QC, missing value analysis, statistical analysis, data mining, machine learning, and pathway analysis.
- **Advanced Statistical Analysis and Machine Learning Application**: Performed visualized high-dimensional clinical data analysis using statistics, machine learning (ML), and data mining methods: 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.
- **Confounding Factor Discovery through Data Mining and EDA**: Through data mining, uncovered a significant confounder that had gone undetected by the research institute for 8 months.

Reason for Job Changes: I had to return to Korea due to unavoidable circumstances: my identification documents were stolen in a home burglary, and subsequent identity renewal became impossible due to the immigration office's indefinite closure during COVID-19.

2018.12 - 2019.05

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

- **Comparative Study and Optimization of Classifiers for Metabolomics Data**: Conducted a comparative study of machine learning methods to evaluate and select the optimal classifier for metabolomics data that best predicts disease states. The methods compared included: 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 Research Competition Award**: Presented a poster at the Annual Research Symposium of Columbia University's Mailman School of Public Health. Selected as one of the top 3 among approximately 100 graduate students in the Annual Research Competition for Master's students, receiving a $1,000 prize and the Chairman's Award.

2014.12 - 2015.06

The City University of New York (CUNY)
Trainee Researcher

- **Mechanistic Modeling**: Developed a mechanistic model reflecting the adsorption process of heavy metals onto tea leaves using differential equations and nonlinear least squares algorithms.
- **Conference Presentations**: Presented a poster at the 2015 Annual Meeting Contributed Papers and Poster Sessions held at New York City College of Technology (CUNY), BMCC (CUNY), and Manhattan College.

2012.08 - 2014.12

Rennert, English Language School in New York City
Trainee Instructor

- Completed comprehensive the ESL (English as a Second Language) coursework
- Underwent TOEFL preparation training
- Earned TESOL (Teaching English to Speakers of Other Languages) certification

- Instructed weekly English classes in English for 30 international volunteer students, covering grammar, reading comprehension, listening, and writing skills.

2010.06 - 2012.02

Molecular Biology Lab, Kangwon National University (KNU)
Trainee Researcher
- Performed protein analysis through cell culture and western blot methodologies.
- Researched the effects and efficacy of Phellinus linteus (Sang-Hwang mushroom) on lymphangiogenesis induced by allergic responses.
- Presented a poster at the Medical Convergence Capstone Design semi-annual symposium and awarded 5 million KRW for outstanding research presentation.

2008.03 - 2010.02

Militarry Service
Military Operations Administrative Coordinator
- Coordinated military strategy education and managed administrative operations
- Managed training documentation and operational planning procedures
- Earned Company Commander's Excellence Award for outstanding performance in division-level ammunition and tactical training inspections
- Achieved early promotion and Division Commander's Recognition Award during Squad Leader Training deployment

Projects

2024.08 - Present

Data Standardization for Data Governance
Seegene, Diagnosis IT General Research Institute
- Role: Project Lead
- Leading enterprise-wide data governance and service integration standardization project
- Scope: 16 departments and 53 databases (Phase 1)
- Approach: Phased data standardization methodology based on existing system analysis and implementation of data quality management framework
- Target Deliverables:
  - 30% improvement in cross-departmental data consistency
  - 30% reduction in data integration time
  - 25% increase in enterprise-wide data utilization
- Technical Skills: Data Governance Framework, Project Management Methodologies, Data Standardization Techniques, Data Modeling, aQueryTool, ETL Processes, Airflow, DBMS

2024.01 - 2024.07

Diagnostic Algorithm Optimization Project
Seegene, Diagnosis IT General Research Institute
- Role: Algorithm Research Scientist
- Standardization and enhancement of sigmoid curve baseline fitting algorithms for improved diagnostic accuracy
- Key Contributions:
  - Analyzed existing methodologies and developed enhanced baseline fitting algorithms
  - Implemented data augmentation and transformation, comparing various techniques (regression models, rule-based algorithms, neural networks)
  - Created intuitive visualizations for non-technical stakeholders
  - Unified various in-house algorithms into a single standardized solution
- Achievements:
  - Improved diagnostic accuracy through optimized algorithms
  - Enhanced algorithm implementation efficiency across the organization
  - Increased stakeholder understanding through clear visual representations
- Technical Skills: Data Analysis, Algorithm Development, Machine Learning, Data

Visualization

2022.12 - 2023.11

FDA Validation Documentation Initiative for Algorithms

Seegene, Diagnosis IT General Research Institute

- Role: Project Manager
- Developed FDA compliance documentation pipeline to facilitate North American market entry and ensure multi-national regulatory compliance
- Key Contributions:
  - Designed and implemented system-level statistical test models for repeatability measurements
  - Led cross-functional team for FDA verification and validation document development
  - Executed data engineering, quality control, and statistical analysis
  - Created comprehensive statistical analysis plans and evaluation metrics
  - Implemented dynamic documentation processes using Quarto, R, and Python
- Achievements:
  - Successfully completed FDA verification and validation documentation
  - Enhanced data quality and reliability of statistical analyses
  - Streamlined documentation process through automated report generation
- Technical Skills: Statistical Analysis, Data Engineering, FDA Regulatory Compliance, Quarto, R Programming, Python, Airflow, Data Visualization

2022.04 - 2022.10

Advanced Diagnostic Algorithm Development

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

- Role: Algorithm Research Scientist
- Developed data-driven signal processing algorithms based on a statistics and sophisticated mechanistic model for diagnostic equipment integrating Seegene's proprietary reagent technology
- Key Contributions:
  - Designed novel mechanistic models incorporating optical characteristics of diagnostic equipment and Seegene's reagent technology
  - Implemented and compared back-fitting and Levenberg-Marquardt algorithms for model optimization
  - Integrated data-driven approaches with mechanistic understanding to enhance algorithm performance
- Achievements:
  - Developed data-driven signal processing algorithms for diagnostics
  - Improved accuracy, interpretability, and reliability of diagnostic results
- Technical Skills: Signal Processing, Mechanistic Modeling, Optimization Algorithms (Back-fitting and Levenberg-Marquardt), Data-Driven Analysis, Scientific Computing

2021.12 - 2022.07

Platform Planning TF

Seegene, Strategy Planning Office & In-house Patent Center

- Role: Strategic Planner / Technical Consultant
- Developed innovative platform strategies and expanded intellectual property portfolio to strengthen Seegene's market position and technical leadership in the diagnostic industry
- Key Contributions:
  - Participated in platform strategy and intellectual property planning initiatives
  - Supported patent claim drafting for selected inventions
  - Provided expert consultation on database systems, statistics, and machine learning to planners and patent attorneys
- Achievements:
  - 16 out of 26 ideas adopted as inventions by patent attorneys
  - 4 patent applications filed with 12 in progress
  - Enhanced collaboration between technical and legal teams

- Technical Skills: Strategic Planning, IP Development, Patent Claim Writing, Database Systems, Statistical Analysis, Machine Learning

2021.01 - 2021.09

Equipment Quality Control Platform Development
Seegene, Diagnosis IT General Research Institute
- Role: Project Lead / Algorithm Developer
- Streamlined and automated medical device quality control processes to improve efficiency and reduce costs and led entire project lifecycle from planning to implementation as project owner
- Key Contributions:
  - Developed enhanced QC algorithms for noise level measurement
  - Created prototype web application using R Shiny to demonstrate automated QC platform feasibility to software engineers
  - Implemented machine learning algorithms for noise test result prediction, significantly reducing QC process time
  - Statistically validated noise test redundancy, enabling process simplification
  - Orchestrated cross-departmental collaboration for QC process automation and visualization
  - Classified various error types: equipment malfunction, human error, reagent production line issues
  - Automated QC platform through web application interface
  - Developed ML-based prediction algorithms for noise test results
- Achievements:
  - Secured 2 patent inventions
  - Reduced annual processing time by 132-fold
  - Decreased annual costs by 13-fold (600 million KRW / $450,000)
  - Eliminated need for separate noise testing in QC process
- Technical Skills: Project Management, Algorithm Development, Machine Learning, Statistical Analysis, Web Application Development (R Shiny), Process Optimization, Cross-functional Collaboration

2020.12 -2024.06

Diagnostic Algorithm Management and Enhancement
Seegene, Diagnosis IT General Research Institute
- Role: Algorithm Developer / Diagnostic Algorithm Manager
- Objectives:
  - Develop and manage diagnostic algorithms reflecting Seegene's proprietary technology and medical device characteristics
  - Implement continuous algorithm improvements based on customer feedback
- Key Contributions:
  - Designed and implemented mechanistic modeling incorporating diagnostic signals and Seegene's proprietary technology characteristics
  - Established signal preprocessing framework using rule-based algorithms
  - Optimized signal processing through Levenberg-Marquardt algorithm-based regression analysis
  - Analyzed VOC (Voice of Customer) feedback for algorithm enhancement
  - Conducted continuous performance monitoring and improvement of diagnostic algorithms
- Achievements:
  - Generated approximately 2.6 trillion KRW ($2 billion) in revenue over 2.5 years through algorithm-applied products
  - Enhanced product reliability through customer feedback-driven algorithm improvements

- Technical Skills: Mechanistic Modeling, Signal Processing, Rule-based Algorithms, Levenberg-Marquardt Optimization, Regression Analysis, Data Analysis, VOC Data Analytics

2018.12 - 2020.04

Long Life Family Study (LLFS) Project

Columbia University Irving Medical Center, Taub Institute

- Role: Research Statistician / Data Scientist
- Investigated metabolomic profiles associated with Alzheimer's disease and brain aging using advanced statistical and machine learning techniques
- Key Contributions:
  - Developed and implemented comprehensive analysis pipeline for metabolomics data
  - Conducted statistical and machine learning analyses to identify Alzheimer's-related metabolomic profiles
  - Performed comparative study of optimal machine learning methods for Alzheimer's and brain aging research
  - Resolved 6-month research bottleneck by identifying critical confounding variables
- Achievements:
  - Identified significant metabolomic profiles associated with Alzheimer's disease
  - Won practical research competition among approximately 100 competitors
  - Received Department Head's Award with $1,000 research scholarship
  - Secured job offer at Columbia University's Taub Institute for Alzheimer's and Brain Aging Research
- Technical Skills: Advanced Statistical Analysis, Machine Learning, Bioinformatics (GWAS), Metabolomics Data Analysis, Research Methodology, Data Analysis Pipeline Development, Missing Data Analysis, Pathway Analysis using Mummichog

2015.01 - 2015.06

Heavy Metal Removal Algorithm Development Using Tea Leaves

The City University of New York, Mathematics

- Role: Research Assistant / Mathematical Modeler
  - Developed mathematical models to predict heavy metal ion removal rates from contaminated water using tea leaves as adsorbents
- Key Contributions:
  - Formulated mechanistic models using differential equations
  - Implemented nonlinear least squares (NLS) algorithms for parameter estimation
  - Conducted experiments to validate model predictions
- Achievements:
  - Presented research at 2015 Annual Meeting's Paper and Poster Sessions across multiple institutions: Manhattan College, New York City College of Technology (CUNY), Borough of Manhattan Community College (CUNY)
  - Awarded $1,000 research stipend
- Technical Skills: Mechanistic Modeling, Differential Equation Modeling, NLS, Optimization Algorithms, Adsorption Kinetics Analysis

2011.01 - 2011.05

Analysis of Phellinus Linteus Effects on Lymphangiogenesis
Kangwon National University, Molecular Biology Laboratory

- Role: Research Assistant
  - Investigated Phellinus linteus (Sangwhang mushroom) as an alternative treatment for allergies, addressing limitations of conventional therapies
- Key Contributions:
  - Conducted experiments to study the impact of Phellinus linteus on lymphatic vessel formation
  - Performed comparative analysis between traditional treatments (corticosteroids and antihistamines) and Phellinus linteus efficacy
  - Executed molecular biology techniques for protein analysis
- Achievements:
  - Presented research findings at Medical Convergence Capstone Design Semi-Annual Symposium
  - Awarded $5,000 research scholarship
- Technical Skills: Cell Culture, Sample Preparation, Protein Isolation, Western Blot Analysis, Detection

Achievements

2023 **Patent Filing (First Inventor)**: Regulatory Model for Diagnostic Algorithms, Seegene
2022 **Patent Filing (First Inventor):** Personalized Treatment Method Based on Iterative Ct Values, Seegene
2022 **Patent Filing (Second Inventor):** Molecular Diagnostics System for Community Groups, Seegene
2022 **Patent Filing (First Inventor):** Subscription System for Medical Platforms, Seegene
2022 **Patent Filing (Second Inventor):** Molecular Diagnostic Negative Certificate, Seegene
2022 **Patent Filing (Second Inventor):** Prediction Model for Molecular Diagnostics, Seegene
2021 **President's Award:** R&D Excellence Award for Noise Test Automation System, Seegene
2021 **Patent Filing (First Inventor):** Automated Noise Test System for Diagnostic Equipment, Seegene
2021 **Patent Filing (First Inventor):** Noise Level Measurement Algorithm for Medical Equipment, Seegene
2021 **Certificate of Completion:** EN62304 Medical Device SW Life Cycle Process Training Course, SGS
2020 **Certificate of Completion:** HIPAA 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, Columbia University
2018 **Certification:** SAS Certified Base Programmer, SAS
2015 **Research Stipend:** $1,000 Mathematical Kinetic Modeling, CUNY
2014 **Certification:** SIT TESOL Instruction Certification, Rennert
2012 **Research Grant:** $5,000 Medical Convergence Capstone Design, Kangwon National University
2012 **Dean's Award:** Summa cum laude, Kangwon National University
2010-2011 **Full Scholarship:** Merit-based Academic Scholarship, Kangwon National University
2009 **Division Commander's Commendation:** Leadership Excellence Award, Military Service
2009 **Company Commander's Commendation:** Division Inspection Excellence Award, Military Service

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