Kwangmin Kim

Data Scientist/ Data Analyst

Email: kmink3225@gmail.com
Website: kmink3225.netlify.app
Linkedin: https://www.linkedin.com/in/kwangmin-kim-a5241b200/

PROFILE

I have over 7 years of work experience in the field of data science, specializing in statistics and machine learning, utilizing open-source tools such as R, Python, SQL, and Apache. I majored in biochemistry, mathematics, and biostatistics, and I am capable of communicating detailed and objective insights derived from modeling and data analysis to non-experts.

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 **Kangwon National University (KNU),** Chun Cheon, South Korea
Biochemistry, Bachelor of Science, *Summa Cum Laude*

ACHIEVEMENT

2023 **Patent Invention,** (main inventor) Validation Model for Algorithm, Seegene (SG)
2022 **Patent Invention,** (main inventor) Medical Intervention by repeatedly measured Cts, SG
2022 **Patent Invention,** (2nd inventor) Diagnostics In-Life Test for Community Group, SG
2022 **Patent Invention,** (main inventor) Subscription System for Medical Platform, SG
2022 **Patent Invention,** (2nd inventor) Molecular Diagnostics Test Result Certification, SG
2022 **Patent Invention,** (2nd inventor) Prediction Model for Molecular Diagnostics Test, SG
2021 **President's Award,** R&D Division Excellence Award for Automatic Noise Test, SG
2021 **Patent Invention,** (main inventor) Automatic Noise Test of Diagnostics Device, SG
2021 **Patent Invention,** (main inventor) Noise Level Measuring Algorithm of Device, SG
2021 **Certificate**, EN62304 - Medical Device SW Life Cycle Process Training Course, SGS
2020 **Certificate**, 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 **Certificate**, SAS Certified Base Programmer, SAS
2015 **Stipend,** $1,000 Mathematical Kinetic Modeling, CUNY
2014 **Certificate,** SIT TESOL Instruction Certification, Rennert
2012 **Stipend,** $5,000 Medical Convergence Capstone Design, KNU
2012 **Dean's Award**, Summa Cum Laude with Academic Excellence Achievement, KNU
2010-2011 **Full Scholarship,** Academic Excellence Achievement, KNU
2009 **Divisional Commander's Award,** Superb Citation in Leadership Competition, Army
2009 **Company Commander's Award,** Superb Citation in Administration Inspection, Army

EXPERIENCE

2020.12 - Present

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

2021 - Present
- Provide consulting services on experimental design, data analysis, and statistical analysis to non-experts, including researchers, strategic planners, executives, and patent attorneys.

2021 - 2024.07
- Managed and optimized diagnostic algorithms for Real-Time PCR signals. To minimize diagnostic errors such as false positives and false negatives caused by abnormal signals and noise, I have progressively improved rule-based algorithms into data-driven algorithms using statistical and mathematical approaches. I optimized the algorithms by statistically analyzing the signal processing and fitting results for each algorithm.

2023
- Planned and authored FDA verification and approval reports for the safety of diagnostic signal processing algorithms using statistical tests. I updated and managed the documentation to comply with FDA regulatory requirements. This project was SG's top priority for 2023, aimed at entering the U.S. market.

2022 - 2023
- Planned and authored the Design History File (DHF) for diagnostic signal processing algorithms. I updated and managed documentation to comply with regulatory requirements from various countries. This project was essential for global business, including Europe, North America, South America, South Africa, and West Asia, and generated approximately 100 billion KRW ($75 million) in quarterly revenue in 2023.

2022 - 2023
- Managed diagnostic algorithms and conducted VOC (Voice of Customer) feedback for processing medical device signal data using the Levenberg-Marquardt algorithm. This diagnostic algorithm helped SG generate approximately 2.6 trillion KRW ($2 billion) in revenue over 2.5 years.

2021.12 - 2022.05
- Was selected for the Platform Strategy IP Planning Task Force Team, where I achieved 26 ideas, 16 inventions, and 5 patent inventions.

2021
- As part of the DevOps project for SG Full Automation, one of Seegene's long-term corporate strategies, I developed a quality control (QC) algorithm for medical equipment, filed two patent inventions, reduced QC process time by over 153 times, and decreased QC costs by 13 times, saving approximately 600 million KRW ($450,000).

2019.05 - 2020.04

Columbia University Irving Medical Center (CUIMC),
Taub Institute for Research on Alzheimer's Disease and the Aging Brain
Research Assistant
- Constructed and suggested an analytic pipeline for the Long Life Family Study (LLFS) using pilot data, including data QC, missing data analysis, statistical analysis, data mining, machine learning, and pathway analysis.
- Performed clinical data analysis with visualization using statistics, machine learning (ML), and data mining: 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.
- Discovered a strong confounder using data mining, which had yet to be identified by the research institute for 8 months.

2018.12 - 2019.05

Columbia University Irving Medical Center (CUIMC),
Taub Institute for Research on Alzheimer's Disease and the Aging Brain
Intern
- Conducted a comparative study of the following machine learning methods to select the optimal classifier for metabolomics data by evaluating the classifier that best predicted the disease status: 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.

- Delivered a poster presentation at the annual research presentation of the Mailman School of Public Health at Columbia University
- Selected as one of the top 3 out of approximately 100 graduate students in the annual research competition for master's graduate students, receiving an award of $1,000 stipend and the Chair's merit.

2014.12 - 2015.06

The City University of New York (CUNY)
Trainee Researcher
- Developed a mechanistic model that reflects the adsorption process of heavy metals into tea leaves using differential equations and a non-linear least squares algorithm.
- Conducted research on certain generalized diffusion models in networks using linear algebra and genetic algorithms (GA).
- Delivered a presentation on the research during the Contributed Paper and Poster Sessions of the 2015 Annual Meeting to be held at Manhattan College, New York City College of Technology (CUNY), and BMCC (CUNY) and received a $1,000 stipend.

2012.08 - 2014.12

Rennert, English Language School in New York City
Trainee Instructor
- Learned English as a Second Language (ESL) Program.
- Studied Test of English as Foreign Language (TOFLE).
- Acquired SIT Teaching English to Speakers of Other Languages (TESOL) Certificate.
- Gave a lecture to 30 volunteer students once a week.

2010.06 - 2012.02

Molecular Biology Lab, Kangwon National University (KNU)
Trainee Researcher
- Conducted quantitative protein analysis using cell culture and western blot techniques.
- Conducted research on the effects of Phellinus Linteus on the formation of lymphatic vessels induced by an allergic reaction, and participated in the experiment to demonstrate its efficacy.
- Delivered a presentation at a semi-annual event of the Medical Convergence Capstone Design, and was awarded a $5,000 stipend.

2008.03 - 2010.02

Military Service
Military Intelligence&Strategy Admin, Squad Commander as a Sergeant
- Worked in Military Strategies Education/Administration.
- Managed administrative tasks and oversaw discipline-planning documents.
- Prepared for ammunition and war game censorship twice a month.
- Received a superb citation in the Leadership Competition and was awarded an early promotion by a divisional commander.
- Received a superb citation in the Administration Inspection from a company commander.

PROJECTS

2024.01 - 2024.07

Baseline Fitting Algorithm 최적화
Seegene, Diagnosis IT General Research Institute
- Analyzed the current state and issues of the existing sigmoid curve baseline fitting and developed an improved baseline fitting algorithm.
- Utilized data augmentation with characteristic equations to improve the baseline fitting method by comparing first-order regression fitting, rule-based processing using first and second-order regression curves, and neural networks. The control group consisted of the existing algorithm and widely-used black-box algorithms in the industry.
- Provided intuitive visualizations to help non-experts, such as researchers and executives, easily understand the improvement results.

- Standardized various non-standardized baseline fitting algorithms within the company into a single unified algorithm.

2023.01 - 2023.11

FDA Verification & Validation Documentation for Diagnostic Algorithms

Seegene, Diagnosis IT General Research Institute

- As a product manager, I am planning a system-level statistical testing model for repeated measurement analysis and writing a statistical analysis plan.
- Implementing the system-level statistical testing model by collaborating with data engineers and biologists, collecting and summarizing documents scattered across departments, and creating evaluation metrics.
- Conducting data engineering, data quality control, and statistical data analysis.
- Writing the FDA verification and validation report documentation for the diagnostic signal processing algorithm using statistical testing as the product manager.

Creating dynamic documentation using Quarto, R, and Python.

2022.07 - 2023.03

DHF Documentation Planning and Writing for Diagnostic Algorithms

Seegene, Diagnosis IT General Research Institute

- Planned and wrote a Design History File (DHF) for a diagnostic signal processing algorithm and its documentation, based on SGS EN62304 and the FDA General Principles of Software Validation document.
- Itemized verification and validation tasks for the diagnostic signal processing algorithm.
- Grouped and formulated the modules of the diagnostic signal processing algorithm for the structural test including unit tests, integration level tests, system level tests, and advanced algorithm tests.
- Created dynamic documentation using Quarto, R, and Python.

2021.12 - 2022.07

Platform Planning TF

Seegene, Strategy Planning Office & In-house Patent Center

- Platform strategy planning and intellectual property planning.
- 16 out of 26 ideas were adopted as inventions by patent attorneys.
- 4 of the 16 inventions have been filed (filing for the rest is ongoing).
- Providing database system, statistics, and machine learning consultation for planners and attorneys.

Data-Driven Diagnostic Algorithm Development

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

- Planned and developed a data-driven signal processing algorithm by defining a newly suggested mechanistic model and applying and comparing the back-fitting algorithm and Levenberg-Marquardt algorithm to the mechanistic model.
- The mechanistic model reflects the optical characteristics of the diagnostic device and the unique technology of SG reagents.

2021.09 - 2021.12

Diagnostic Device Quality Control (QC) Platform Construction

Seegene, Diagnosis IT General Research Institute

- Led all processes throughout the project as the project owner.
- Developed an improved QC algorithm measuring a noise level.
- Collaborated with other departments to automate the QC process and visualize the QC process.
- Classified device failures, human errors, and reagent production line errors.
- Developed a web application of an automatic QC platform as a prototype to demonstrate the project's feasibility to software engineers.
- Developed an algorithm to predict a noise test result as the second stage of a QC process on calibration data as the first stage using machine learning, to reduce the

time-consuming QC process.
- Statistically demonstrated that the time-consuming noise test with medical devices is not necessary in the QC process, which led to the simplified QC process.
- Achieved the 2 patent inventions, the abolition of the noise test, reduced turnaround time by >132x a year and about \600 million ($450,000) cost by 13x a year in the QC process.

2018.12 - 2020.04

Long Life Family Study (LLFS) Project
Columbia University Irving Medical Center, Taub Institute
- Conducted statistical and machine learning (ML) analysis to identify metabolic profiles significantly associated with Alzheimer's Disease.
- Constructed an analytics pipeline including missing value analysis, statistical analysis, ML classification, and pathway analysis using the Mummichog tool.
- Conducted a comparative study of optimal machine learning (ML) methods for Alzheimer's Disease and the Aging Brain using metabolomics data.

2015.01 - 2015.06

Heavy Metal Removal Algorithm Development Using Tea Leaves
The City University of New York, Mathematics
- Researched for modeling adsorption kinetics with differential equations and non-linear least square algorithm: With a hypothesized mechanistic model of the adsorption process, demonstrated the model is useful for predicting the rate at which tea leaves can remove heavy metal ions from polluted water.

Generalized Diffusion Model in Networks using Linear Algebra and GA
The City University of New York, Mathematics
- Researched for theoretical generalized diffusion modeling in networks using methods from linear algebra to highly connected networks in order to examine multi-layered information exchange using genetic algorithms.

2011.01 - 2011.05

Effects of Phellinus Linteus toward Formation of Lymphatic Vessel
Kangwon National University, Molecular Biology Lab
- Researched for effects of Phellinus Linteus toward the formation of lymphatic vessels Induced by allergic reaction: although allergic reactions have been treated with medicines of the kinds of corticosteroid and anti-histaminic agent and many other kinds of medicines have been developed, their efficacy is temporary. Accordingly, Phellinus linteus can play an alternative role in treating allergies.

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), Streamlit (beginner)
Engineering	● Airflow (intermediate)
SQL	● PosgreSQL, 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