Kwangmin Kim

Data Scientist/ Data Analyst

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

PROFILE

Over 7 years of experience in data science with academic backgrounds in biochemistry, mathematics, and biostatistics and expertise in statistics and machine learning using open source tools such as R, Python, SQL, etc. I am interested in gaining a mathematical understanding of algorithms and modeling. I can communicate with non-experts using detailed objective facts obtained through data analysis. I aim to work in a planned, systematic way.

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

- Planned and wrote a Design History File (DHF) for a diagnosis signal processing algorithm. Currently updating and managing the documentation to ensure compliance with regulatory requirements. This project enabled global business including Europe, North America, South America, Southern Africa, and West Asia, Sales of about ₩100 billion ($75 million) per quarter.
- Planned and wrote the FDA verification and validation report documentation for the safety of the diagnostic signal processing algorithm using statistical testing, resulting in 1 patent invention. Currently updating and managing the documentation to ensure compliance with regulatory requirements. The project is SG's top priority project to enter the US market.
- Manage a diagnosis algorithm for processing signal data from a medical device, utilizing the Levenberg-Marquardt algorithm. With the diagnostic algorithm, SG recorded sales of approximately ₩2.6 trillion ($2 billion) over a period of 2.5 years.
- Provided consulting on experimental design, data analysis, and statistical analysis to non-experts, such as experimenters, strategy planners, executives, and attorneys.
- Participated in the Platform Strategy IP (Intellectual Property) Planning TF, and achieved 26 ideations, 16 inventions, and 5 patented inventions.
- Developed Quality Control (QC) algorithms for device, resulting in filing 2 patented inventions, a >153x reduction in time spent in the QC process, and a 13x reduction amounting to ₩600 million ($450,000) in QC costs.

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 not been 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 rorests, 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

2022.07 - present

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.

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.

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	Database	etc.
• R, Python	• SQlite	• Ubuntu, Powershell, Git/Github, Conda
• SAS	• Oracle-SQL	• Quarto, R markdown, Jupyter,