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

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

ACHIEVEMENT

2022 **Patent Invention,** Customized Intervention by Time-series Analysis of Ct Values, Seegene
2022 **Patent Invention,** Molecular Diagnostics In-Life Test for Community Group, Seegene
2022 **Patent Invention,** Subscription System for Medical Platform, Seegene
2022 **Patent Invention,** Molecular Diagnostics Test Result Certification, Seegene
2022 **Patent Invention,** Prediction Model for Molecular Diagnostics Test, Seegene
2021 **President's Award,** R&D Division Excellence Award for Automatic Noise Test, Seegene
2021 **Patent Invention,** Automatic Noise Test of Diagnostics Device, Seegene
2021 **Patent Invention,** Noise Level Measurement Algorithm of Medical Device, Seegene
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
- Planning a Design History File (DHF) for a diagnosis signal processing algorithm.
- Planning and writing the FDA verification and validation report documentation for the safety of the diagnostic signal processing algorithm using statistical testing.
- Management of diagnosis algorithms processing signal data from a medical device.
- Device Quality Control (QC) algorithms development.
- Consulting on data analysis and statistical analysis to non-experts.
- IP (Intellectual Property) planning and filing 7 patent inventions.

2019.05 - 2020.04

Columbia University Irving Medical Center (CUIMC),
Taub Institute for Research on Alzheimer's Disease and the Aging Brain
Research Assistant
- Clinical data analytics with visualization using statistics and machine learning (ML).
- Pipeline construction: data QC, missing data analysis, statistical analysis, data mining, machine learning, and pathway analysis.
- ML and statistics consulting to medical doctors, epidemiologists, and neurologists.
- Gave a poster presentation in the annual research presentation of the Mailman School of Public Health at Columbia University.

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 of machine learning methods for selecting the optimal classifier for metabolics data.
- Performed dimension reduction for high dimensional data to deal with highly correlated variables: variable extraction and selection using Lasso, ridge regression, elastic net, principal component analysis,, partial least square, and sparse-partial least square.
- Gave a poster presentation in the annual research presentation of the Mailman School of Public Health at Columbia University.

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 equation and non-linear least square algorithm.
- Researched certain generalized diffusion models in networks using linear algebra and genetic algorithms.
- The CUNY Research Scholar Program awarded a $1,000 stipend.
- Gave a presentation on the research for the Contributed Paper and Poster Sessions of its 2015 Annual Meeting to be held at Manhattan College, at New York City College of Technology (CUNY), and at BMCC (CUNY).

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 the western blot.
- Researched for effects of Phellinus Linteus toward the formation of lymphatic vessels induced by an allergic reaction and joined the experiment to demonstrate its efficacy.
- Gave a presentation for a semi-annual event of Medical Convergence Capstone Design.

2008.03 - 2010.02

Military Service
Military Intelligence&Strategy Admin, Squad Commander as a Sergeant
- Worked on Military Strategies Education/ Administration.
- Took charge of administrative tasks and managed discipline-planning documents.
- Prepared for ammunition and scheme of war game censorships twice a month.

PROJECTS

2022.07 - present

DHF Documentation Planning and Writing for Algorithms
Seegene, Diagnosis IT General Research Institute
- Planning and writing a Design History File (DHF) for a diagnosis signal processing algorithm and its documentation based on SGS EN62304 and the FDA General Principles of Software Validation document.
- Dynamic documentation using Quarto, R, and python.

FDA Verification & Validation Documentation for Algorithms
Seegene, Diagnosis IT General Research Institute
- As a product manager, planning a system-level statistical testing model and writing a statistical analysis plan.
- Implementing the system-level statistical testing model by collaborating data engineers and biologists.
- Conducting 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.
- 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 were filed (filing for the rest is ongoing).
- Database system, statistics, and ML 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.
- Reflected optical characteristics of the diagnosis device.
- Reflected the unique technology of SG reagents.

2021.09 - 2021.12

COVID-19 Clinical Data Analytics
Seegene, Diagnosis IT General Research Institute
- Conducted data quality control and demographic analysis of COVID-19.
- Conducted quality control of the diagnostic device and assay reagent product.
- Developed an algorithm for detecting a mutation suspect signal.

2021.01 - 2021.09

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, and reduced turnaround time by 11x.
- 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.
- Applied for a patent in recognition of the patentability of this invention.

Noise Test Result Prediction

Seegene, Diagnosis IT General Research Institute

- 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.
- Applied for a patent in recognition of the patentability of this invention.

2019.05 - 2020.04

Long Life Family Study (LLFS) Project

Columbia University Irving Medical Center, Taub Institute

- Conducted statistical and ML analysis to identify metabolic profiles significantly associated with Alzheimer's Disease.
- Analytics pipeline construction: missing value analysis, statistical analysis, ML classification, and pathway analysis using the Mummichog tool.
- Using data mining, a strong confounder was discovered that had not been found by the research institute for 8 months.

2018.12 - 2019.05

Analytic Project on Alzheimer's Disease and the Aging Brain

Columbia University in the City of New York, Biostatistics

- Comparative study of optimal ML methods for Alzheimer's Disease and the Aging Brain and metabolomics.
- Selected as one of the top 3 out of about 100 graduate students in the annual research competition for master’s graduate students.

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.
- 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 the genetic algorithm.
- The CUNY Research Scholar Program awarded me a $1,000 stipend.

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,