Kwangmin Kim

Data Scientist/ Data Analyst

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

ABOUT ME

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.

SKILLS

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

EXPERIENCE

2020.12 - Present

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

- Planned and authored 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 authored 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 devices, resulting in the achievement of 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:
  - 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, which resulted in the abolition of the noise test.

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.

EDUCATION

2017.08 - 2019.05 ● Columbia University in the City of New York, Biostatistics(MS), Department Head Award
2015.08 - 2017.05 ● Baruch College, The City University of New York, Mathematics(BA)
2006.03 - 2012.02 ● Kangwon National University, Biochemistry(BS), Summa Cum Laude, Full Scholarship