Yeonji Ji, Ph.D.

Cliffside Park, NJ | fortunate.yj@gmail.com | 929-810-5531 | LinkedIn | GitHub | Portfolio

Summary

Data Scientist with a Ph.D. and experience in **statistical modeling, large-scale data analysis, and machine learning** (Python, SQL, PyTorch). Skilled in building scalable pipelines and delivering insights that drive product decisions.

Technical Skills

Programming:
Python (pandas, scikit-learn, PyTorch, matplotlib, seaborn, etc.), Bash, Git, C++

Machine Learning:
Supervised/unsupervised learning, Feature Engineering, Hyperparameter Tuning, Model Evaluation, Ensembles, Deep Learning

Computing:
Linux, HPC (Slurm), Workflow Automation

Specialized Expertise:
Computational Drug Discovery, Molecular Simulations, Virtual Screening, Molecular Visualization

Education

Ph.D. in Biochemistry
The Graduate Center, CUNY *2018 – 2024*
Dissertation Title:
"Incorporating Solvation Thermodynamic Mapping in Computer-Aided Drug Design"

B.S. in Chemistry
Kyung Hee University, Seoul *2012 – 2017*

Certifications

Python for Machine Learning and Data Science Masterclass (Udemy)
Supervised/unsupervised ML, PCA, model evaluation

Deep Learning Specialization (Coursera)
Neural networks (CNN, RNN, LSTM, etc.), optimization, project structuring

Honors & Grants

CUNY DSRG (2023)
Penny J. Gilmer Grant, OpenEye (2023)
CUNY Science Scholarship (2018–2024)
Superiority Scholarship, KHU (2014–2015)

Machine Learning Projects

Amazon Review Sentiment Classification (GitHub) *2025*
- Built a text classification model using TF-IDF and ML algorithms (Logistic Regression, XGBoost).
- Conducted feature engineering, hyperparameter tuning, and evaluation on 3M+ reviews.

Teleco Customer Churn Prediction (GitHub) *2025*
- Developed churn prediction models using Logistic Regression, Random Forest, and XGBoost.
- Applied SMOTE and feature engineering on 7k+ customer records to address class imbalance and improve prediction accuracy.

Research & Data Projects

Binding Site Prediction from Simulation Data *(Publication)* *2023 – 2025*
- Processed large-scale molecular dynamics datasets, analyzing time-series trajectories and 3D solvation data to identify cryptic pockets.
- Used statistical models to quantify solvation energy costs and assess pocket binding potential.

Data-driven Pharmacophore Modeling *(Publication in process)* *2021 – 2024*
- Engineered automated Python workflows to extract hydration features and construct water-derived pharmacophore models.
- Evaluated performance on large compound libraries, enhancing screening reproducibility and scalability.

COVID-19 Solvation Mapping Repository *(Publication)* *2019 – 2020*
- Contributed analysis tools and datasets to an open-source GitHub repository, leading to a peer-reviewed publication.

Professional Experience

Postdoctoral Researcher *Lehman College, CUNY* *2024 – Present*
- Leading independent research on binding site thermodynamics and molecular simulations.
- Developing Python workflows and preparing manuscripts; mentoring graduate students and collaborating across the lab.

Adjunct Lecturer *CUNY Research Foundation* *2019 – 2024*
- Taught General Chemistry Lab to 100+ undergraduates, simplifying complex concepts through hands-on instruction.

Cosmetic Chemist Intern *Englewood Lab, NJ* *2015 – 2016*
- Conducted R&D experiments to optimize cosmetic formulations, producing samples that met stability and performance standards.