Ji In Lee (Erin)

(510) 990-5227 | erinlee@berkeley.edu | linkedin.com/in/erin-ji-in-lee | github.com/erinlee99/datascience

EDUCATION

University of California, Berkeley Aug 2017 – May 2022
Bachelor's Degree in **Data Science** *Merit-based Cal Alumni Association Scholarship*

Coursework: Principles of Data Science; Computational Structures in Data Science; Data Mining and Analytics; Machine Learning; Data Structures; Probability for Data Science; Natural Language Processing; Linear Algebra; Statistics; Economics

SKILLS

- SQL, Python (pytorch, scikit-learn, numpy, pandas, matplotlib, seaborn), JAVA
- Tableau, Excel, Adobe Analytics, Google Analytics, AWS, Git, Jupyter Notebooks, SAS, Kibana, Microsoft Excel

WORK EXPERIENCE

EY (Ernst & Young) July 2020 – Sept 2020
Data and Analytics Intern *Received offer to extend original contract*

- Diversified client's user base by analyzing and modeling on 21.5M+ instances of raw user data from a Big Data platform utilizing SQL & SAS and translating results into customer segmentation strategies for a leading national bank.
- Collaborated with a team of 6 consultants to optimize Customer Relationship Management Database and lead Persona Workshops to develop hyper-personalized products for the client's existing and targeted user groups.
- Developed a business proposal for the first South Korean IoT smart hospital at a healthcare startup, 30+ companies benchmarked.

INNODS INC. Nov 2021 – Feb 2022
Data Scientist Intern

- Developed Tableau dashboards and data visualizations of sales prediction models for businesses on *Naver Smartstore*, Korea's top sales platform. Received the National Financial Commission Chairman Award (1st place).
- Proposed a credit score model utilizing logistic regression with 11.7% higher accuracy, compared to the team's methodology, by analyzing key variables correlated to 15.8M+ online sales data.

Lawrence Berkeley National Laboratory Sept 2020 – May 2021
Undergraduate Research Assistant

- Ran Data Pipelines on neuronal models and wrote technical documentation (ReadMe) for plotting neuronal functions.
- Researched about optimization algorithms to model biophysically detailed models of ion channel distribution using Python.

Resiliency Technologies Aug 2021 – Dec 2021
Data Scientist Intern

- Took ownership of client deliverables, creating data visualizations that analyzed behavior patterns from raw user data on a mental health database. Utilized Google Analytics and Kibana.
- Converted the existing library of modules into *SCORM* format using JSON, HTML with the goal of increasing user engagement.

INNODS INC. Dec 2018 – Jan 2019
Data Analyst Intern

- Designed Kibana dashboards with 10+ data visualizations of real-time flight reservation data, translated into marketing strategies.

SELECTED PROJECTS

H&M Personalized Recommendations *UC Berkeley x Seoul National University : Team Leader* March 2022 – May 2022

- Developed a K-Means cluster recommendation algorithm for 13.6M+ customers using AWS, Python, customer/product meta data
- Improved model performance by 34% by implementing a decoding algorithm that quantified the similarity between consumption patterns of encoded postal codes using vectorization, isomorphism, clustering.

Classifier Identifying Fraudulent Job Postings *Team Co-Leader* Dec 2021

- Implemented a classifier that identifies fraudulence from 1.7M+ online job posting data using Random Forest, Logistic Regression
- Extracted determining features that correctly identify fraudulence by implementing a logarithm dictionary algorithm