Edward Sun

itsedwardsun@gmail.com | 408-772-9629 | LinkedIn | GitHub

4 years of data science experience. Proficient in translating complex analytical and technical concepts to non-technical employees, initiating and driving projects to completion with minimal guidance.

Education

University of Chicago | Chicago, IL — Sep 2020 – Aug 2021
M.S. Biomedical Informatics. **GPA:** 3.92

University of California, Berkeley | Berkeley, CA — Aug 2016 – May 2020
B.A. Data Science. Chemical Engineering Minor.
Coursework: Database Systems (relational databases and SQL), Probability Theory, Efficient Algorithms and Intractable Problems, Principles of Data Science

Work Experience

CareDx | Brisbane, CA — March 2022 – Present
Data Scientist

- Proposed and implemented a gradient boosted tree model with Bayesian optimization and Spark to support AlloSure, a clinically validated test for transplant patients. Obtained a model with 0.97 AUC using 1 million patient labels
- Collaborated with team of 9 peers and cross-functionally with R&D and IT teams to optimize product changes
- Built machine learning models in Databricks to predict transplant rejection events for >30,000 patients

AbbVie | Chicago, IL — January 2021 – August 2021
Machine Learning Intern

- Generated predictive machine learning model for >17,000 donors to predict genotypic data at 10,000 positions
- Optimized ensemble classifier and attribute bagging model in R to achieve >95% prediction accuracies
- Produced ancestry-specific databases applicable to both common and low-frequency variants

Industrial Technology Research Institute | Zhudong, Taiwan — June 2019 – Aug 2019
Deep Learning Intern

- Developed a convolutional neural network for image classification using over 100k video frames
- Converted computer vision research papers based on AlexNet into working C++ code
- Achieved 95% accuracy on class-labeled video image frames with Caffe and CUDA

Skillset

Programming*:* Java, Python (Numpy, Pandas, Scikit-learn, Jupyter notebooks, Matplotlib), SQL, R, Git, MATLAB.
Data Science*:* Docker, Bioconductor, Tableau, Apache Spark.
Modeling*:* Supervised Learning (linear and logistic regressions, decision trees, support vector machines), Unsupervised Learning (k-means clustering, principal component analysis), CNNs

Projects

Predictive Model for Cervical Cancer Risk Factors — June 2021

- Pre-processed cervical cancer data from UCI Machine Learning Repository to determine presence of cervical cancer using patient factors such as health history and lifestyle
- Utilized logistic model with ridge regression and feature engineering to achieve an AUC of 0.89

Kaggle Competition Email Classification — March 2020

- Designed a binary classifier to distinguish spam emails using text mining, feature engineering, and natural language processing (NLP). Achieved top 4% (36/864) in test accuracy