Christopher Bui

Email: cbui3@pm.me **Phone:** 714-334-2810 **Blog:** christopherbui.github.io **LinkedIn:** /in/cbui3 **GitHub:** GitHub

Education

Johns Hopkins University
Bioinformatics, M.S. *(Graduation: 2022)*

University of California, San Diego
Biophysics, B.S.; Entrepreneurship Minor *(Graduation: 2018)*

Experience

Data Scientist — Long Beach, CA
SCAN Health Plan — Nov. 2023 – Current

- Developed efficient SQL queries on Microsoft Azure platform to analyze patient healthcare data
- Led the research and development of multiple CART and Deep Neural Network classifiers using PyTorch to identify patients at high risk for medication non-adherence
- Created classifier model and assessed performance in determining member's eligibility for healthcare plans

Data Scientist — Irvine, CA
Bio-Rad Laboratories — Sep. 2023 – Oct. 2023

- Developed efficient SQL queries in Snowflake environment to optimize datasets for PowerBI dashboards
- Analyzed profit margins of company's entire product portfolio to identify key performance indicators
- Built and presented professional dashboards using Power Bi and associated DAX language

Data Scientist — Laurel, MD
The Johns Hopkins Applied Physics Laboratory — Aug. 2022 – Jul. 2023

- Built data preprocessing pipeline in Python for hospital operating room time series sensor data
- Deployed supervised and unsupervised models including linear regression, random forest, and clustering to predict
- Utilized High Performance Computing (HPC) clusters through SLURM on Linux to run Sars-Cov-2 protein-protein docking simulations
- Used *NextStrain* open-source tool to compare Sars-Cov-2 strains
- Analyzed federal inventory of Covid-19 personal protective equipment using PySpark
- Wrote and executed standard operating procedure (SOP) for field testing hardware that predict real-time core body temperatures of military working dogs

Machine Learning Engineer Intern — Silver Spring, MD
3M — Jun. 2020 – Aug. 2020

- Utilized the PySpark library on AWS to process parquet text files containing medical records data
- Designed machine learning models to create text embeddings for ICD-10 medical codes to improve 3M's NLP software capabilities
- Followed a Scrum framework and managed tasks using Jira

Projects

Cancer Tissue Detection

- Developed a PyTorch CNN model and applied transfer learning to classify tissue images as either cancerous or not
- Trained on 200,000 images using CUDA tensor cores; Performance reached 86% overall accuracy

Semantic Image Segmentation of Cellular Nuclei

- Created a U-Net architecture neural network using Tensorflow and trained model on AWS to segment nuclei locations
- Preprocessed 700 microscope images of nuclei of varying cell types using Scikit-learn and OpenCV

Lymphoma Microarray Analysis

- Applied Principal Component Analysis on high dimensional lymphoma cancer microarray dataset
- Utilized k-means and hierarchical clustering techniques to classify lymphoma strains and group similar functioning genes

Wikipedia Question & Answer

- Created an unsupervised NLP model using TF-IDF and cosine similarity that outputs relevant text in response to an input question
- Utilized NLTK and Gensim to preprocess JSON text data of Wikipedia articles

Skills

Technologies: Python, R, Java, PyTorch, Tensorflow, Keras, Scikit-Learn, Pandas, Numpy, Jupyter Notebook, PySpark

Data Management: PostgreSQL, Git, AWS, Spark, Tableau, PowerBI, Microsoft Azure

Soft: Verbal & Written Communication, Detail-Oriented, Self-Motivated, Organized, Empathy