Christopher Bui

Email cbui3@pm.me | **Phone** 714-334-2810 | **Blog** www.cbui.me | **LinkedIn** /in/cbui3 | **GitHub** christopherbui

Education

Johns Hopkins University Bioinformatics, M.S. (2022)

University of California, San Diego Biophysics, B.S., Entrepreneurship, Minor (2018)

Experience

Analytics Engineer Boston, MA

Pure Lithium Corporation Dec. 2024 – Present

- Optimized SQL queries using Snowflake and Data Build Tool (DBT) to generate data tables for dashboards on Hex for stakeholders.

Data Scientist Long Beach, CA

SCAN Health Plan (Remote) Nov. 2023 – May. 2024

- Developed optimized SQL queries on the Microsoft Azure Data Platform to analyze patient EHR and claims data, enabling insights into healthcare trends and client behavior.
- Led development of knowledge graph embedding and CART models, and Deep Neural Network classifiers using Scikit-Learn and PyTorch, to predict high-risk patients for medication non-adherence with improved precision.
- Deployed machine learning models in production, evaluating their performance to enhance the identification of members eligible for healthcare plans through AUC-ROC metrics.

Clinical Programming Analyst Irvine, CA

Edwards Lifesciences (Contract) Oct. 2023 - Nov. 2023

- Ensured consistent schema of KPI dataset through appropriate data validation and preprocessing using PySpark.
- Optimized interactive data visualizations by enhancing Python Dash Plotly dashboards, improving usability for clinical stakeholders.

Data Scientist Irvine, CA

Bio-Rad Laboratories Sep. 2023 – Oct. 2023

- Developed optimized SQL queries in Snowflake to accelerate dataset preparation for PowerBI dashboards built using DAX language.
- Conducted profit margin analysis across the company's product portfolio to identify actionable KPIs, improving product strategy.

Data Scientist Laurel, MD

The Johns Hopkins University Applied Physics Laboratory Aug. 2022 – Jul. 2023

- Built a scalable ETL data pipeline in Python to preprocess environmental sensor data from hospital operating rooms.
- Developed regression and clustering models to predict hazardous operating room conditions, improving patient safety protocols.
- Conducted hyperparameter tuning of KGE models including ComplEX and TransE for SARS-Cov-2 target protein identification.
- Leveraged High Performance Computing (HPC) clusters with SLURM on Linux to run SARS-CoV-2 protein-protein docking simulations.
- Analyzed Covid-19 personal protective equipment inventory using PySpark, identifying supply chain bottlenecks.
- Created Standard Operating Procedures (SOPs) for field-testing IoT hardware that predicted core body temperatures in real-time.

Machine Learning Engineer Intern Silver Spring, MD

3M Jun. 2020 – Aug. 2020

- Leveraged PySpark and NTLK on AWS to preprocess large-scale parquet text files containing medical records.
- Designed NLP models to generate text embeddings of ICD-10 medical codes, enhancing medical code recommendation accuracy.

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

Next Generation Sequencing Analysis Pipeline

- Automated a standard NGS sequencing analysis pipeline using Python and SRA data to generate basic statistics and identify variants

Skills

Technologies: Python, R, Java, HTML/CSS/Javascript, Scikit-Learn, Pandas, Numpy, Tensorflow, PyTorch, PyKeen, Optuna, Jupyter Notebook, OpenCV, SQL, PySpark, Git, Amazon Web Services, Microsoft Azure, Snowflake, HPC, SLURM, Tableau, PowerBI

Knowledgeable: Predictive Modeling, Statistical Analysis, Deep Learning, Knowledge Graph Embeddings, Natural Language Processing, Time Series Forecasting, Computer Vision, A/B Testing, Feature Engineering, Data Visualization, Communication, Self-Motivation, Empathy