Daniel Chen

Data Scientist | Machine Learning Engineer | Software Engineer

Millbrae, CA, USA
www.linkedin.com/in/dchen71
www.github.com/dchen71
www.thisisdchen71.com

EXPERIENCE

Senior Bioinformatics Associate 1 — *Gilead Sciences Inc*
11/2018 - PRESENT

● Architect and collaboratively build automated Kubernetes Airflow deployment for enterprise level ETL jobs in order to save 50-80% in cloud costs by going cloud native for analysis

● Manage vendor relationships and provide technical input for rapid POCs into potential technology acquisitions for digital transformation of company

● Support development of image recognition model using UNET model to classify cells for potential clinical trial usage

● Statistically model and engineer framework to classify risk of any R packages for regulatory compliance and cost savings from potential technology partnership

● Develop and maintain full stack data applications to allow nontechnical people access to visualizations and results

Biostatistician 1 — *Gilead Sciences Inc(Akraya Inc)*
11/17 - 11/18

● Analyzed next generation sequencing data for the selection of sequencing technology vendor and replacement of contract research organization

● Created analytical applets and workflows using the DNAnexus platform and pioneered reproducible Docker container based workflows to save development time and costs

● Developed early data warehouse for genomic data and querying API to transition towards database technologies for visualization and analysis workflows

● Collaborated cross functionally to create full stack web applications for visualization of data

Scientist 1 — *Roche Sequencing Solutions (Aerotek Inc)*
04/17 - 11/17

● Designed and executed experiments as well as analyzed novel proteins to find relevant sequencing characteristics

● Developed and maintained web applications to make data accessible cross functionally to allow non-programmers access to sequencing data visualization and analyses

● Developed automation bash scripts to simplify pre-sequencing steps in experiments and to save time per sequencing run

● Implemented and designed improved SOP for experiments to preemptively deal with potential sequencing failures due to wafer defects

● Optimized existing ETLs from days to seconds, freeing computational resources for downstream analysis

● Built postgresql data warehouse to simplify and centralize data for data analysis and visualization

SKILLS

Languages
Python, R, Bash, Javascript, SQL
Data Engineering
Docker, postgresql, mySQL, Oracle, AWS, Kubernetes, Apache Airflow, Microsoft Azure, node.js, Express, Ansible, Git, dask, spark
Data Visualization
Shiny, plotly, Tableau, d3.js, react.js, vue, nuxt, django
Data Analysis
tidyverse, pandas, spacy, numpy
Machine Learning
Tensorflow, scikit-learn, NLP, Reinforcement Learning, Deep Learning

EDUCATION

MBA — *University of Southern California*
08/2021 - 05/2023

ALM Data Science — *Harvard University Extension School*
06/2020 - 05/2021

BS Biotechnology— *University of California, Davis*
09/2008 - 12/2011

PROJECTS

Scientific Text Lay Summarization

Building abstractive text summarization models to simplify scientific text for general audience

Real Time Mask Detection

Utilize YoloNetV4 for real time mask detection using webcam