QUOC-HUY NGUYEN
(714) 837-0205 | quochuyn1999@gmail.com | linkedin.com/in/quochuyn | quochuyn.github.io

EDUCATION

University of Michigan May 2022 - Aug 2023
Master of Applied Data Science 3.9 GPA
University of California, Irvine Sep 2017 - Jun 2021
B.S., Mathematics - Data Science Specialization; Statistics Minor Summa Cum Laude, 3.9 GPA

EXPERIENCE

Upwork Remote
Data Scientist (Freelancer) May 2022 - Present

- Wrote Python using TensorFlow to evaluate an overfitting convolutional neural network (CNN) for a client researching medical imaging in healthcare.
- Coded in Python using GeoPandas to provide analytical insights and geographic visualizations on malicious hacking attempts.
- Wrote SQL queries to analyze customer billing transactions and presented insights and code to a stakeholder.

Curacao Los Angeles, CA
Data Analyst Sep 2021 - Apr 2022

- Engineered a data pipeline for our mobile application product, enabling the extraction, transformation, and loading (ETL) of 16+ GB of Open Banking data from FinTech APIs to underwrite personal loans.
- Designed and developed a scikit-learn random forest classification model that predicts loan payment with more than 70% accuracy.
- Analyzed business data to inform and recommend our marketing strategy across the Customer Lifecycle.
- Optimized routine reports in SAS and Excel that resulted in saving about 10% of time spent.
- Acquired proficiency in SQL and SAS within 2 months, leveraging these tools to execute data analytic tasks.

ACADEMIC DATA SCIENCE PROJECTS

Youtube NLP Recommender June 2023 - Present

- Developed an end-to-end solution that integrated with the Youtube API and a website for real-time content viewing and filtering with Natural Language Processing (NLP) techniques.
- Configured an AWS RDS PostgreSQL database for user preferences that integrated with the Streamlit web app.

NHL Draft Predictions May 2023 - June 2023

- Created a machine learning model to predict NHL draft positions for prospective hockey players and cluster players together into groups based on the nearest neighbor algorithm.
- Harnessed the cutting-edge power of Language Models (LLMs) including Google's BERT and OpenAI's Chat-GPT to encode word embeddings as model features.

RESEARCH

Atwood Lab Irvine, CA
Undergraduate Researcher Aug 2019 - Mar 2021

- Researched skin cancer single-cell data using methods for principal graphs and graph neural networks.
- Developed a graph-based semi-supervised classification Pytorch model for identifying cancerous cells.

MathBioU & Math EXPLR Irvine, CA
Undergraduate Researcher Jul 2019 - Aug 2019

- Performed computational biology research at a 6-week summer research program.
- Collaborated as a co-author in a research report to present findings to a diverse >15 person audience.

SKILLS

Programming: Python (4 years), SQL (2 years), PostgreSQL, R, SAS, C++, Matlab, Bash, Shell, Unix/Linux Terminal
Libraries: Numpy, Pandas, Scikit-learn, PyTorch, TensorFlow, Hugging Face, PySpark, Statsmodels, NLTK, Matplotlib, Seaborn, Altair, Geopandas, Psycopg2, Streamlit
Datasets: Customer, Financial, B2B, Healthcare/Clinical, Geospatial, Causal Inference, JSON, Experiments
Skills: Machine Learning, Neural Networks, NLP, Version Control (Git, Github, DVC), Jupyter Notebook, Business Intelligence (Power BI, Domo), Microsoft Excel, Cloud Infrastructure (AWS S3/EC2/RDS), DBeaver, RStudio, LaTeX