Minh Tran

Phone Number: 703-626-8840
Email: Minhtraann@yahoo.com
LinkedIn: https://www.linkedin.com/in/minh-tran-a5206616a/
Github: https://github.com/Minhereno

ABOUT

A data analyst/data scientist open to positions to learn and utilize my knowledge and skills in exploratory data analysis, data manipulation, AI/ML, and data visualization to turn into actionable insights to benefit the public good (Python, R, Microsoft Excel, SQL, and Tableau). I have experience working in an AGILE environment in a federal contractor setting and worked with machine learning and real world statistical metadata such as image/document/videos/csv/xlsx formats as well as public health metadata and FASTA/FASTQ data formats.

SKILLSET

- Computational Data Science:
  - Python (Numpy, Pandas, Matplotlib, Scikit-learn, Pytorch, Tensorflow, spaCy, CUDA)
  - R (dplyr, ggplot, tidyverse)
  - SQL
  - Tableau
  - Microsoft Office/Microsoft Excel (Pivot Tables, Lookup functions, Macro Functions)
- Linux, VMware
- Github & Bitbucket (Version Control)
- Docker (Application Export)
- JIRA/Confluence (AGILE Methodology)

EDUCATION

George Mason University — **Fairfax, VA**
Bachelor of Science in Bioinformatics — *Aug 2017-May 2021*

EXPERIENCE

DATA ANALYST/DATA SCIENTIST — **Reston, VA**
Sept 2021 – Present

NOBLIS (Full Time)

- Led efforts to develop and document pipeline combining open-source and state-of-the-art large language models with NLP to extract relevant information from documents/images (Python)
- Took over handoff work to clean and aggregate large Influenza dataset with more than 2 million data points with pivot tables and report on key metrics. Trained and evaluated classifier model to predict Influenza subtypes and reported key performance metrics based on Meta's ESM2 sequence embeddings (Precision, F1-score). (Microsoft Excel, Python)
- Contributed and documented on implementing code changes to train/evaluate image/video classification/clustering machine learning models to achieve greater than 0.90 F1-Score (Python)
  - Data cleaning, formatting, and analysis (Microsoft Excel, CSV, JSON, XLSX)
- Worked on isolating different AI uses cases of pipeline into individual microservices (Docker, Kafka)
  - Reduced build time and image size of microservice server environment by more than 30%
- Extracted and analyzed metadata and metrics of SARS-CoV-2 genomes (FASTA/FASTQ) from public repositories (Python, Microsoft Excel)
  - Created Python pipeline to automate processing COVID genomes and submit to public repositories such as NCBI
  - Created comprehensive dashboard visualizing the metadata of SARS-CoV-2 with more than 1 million data points (Microsoft Excel, Tableau)

BIOINFORMATICS INTERN **Reston, VA**

Jan 2020 – Aug 2021

NOBLIS (Part Time)

- Completed data scraping, statistical analysis, and visualization of SARS-CoV-2 testing metrics (Number of people infected/hospitalized by county in Virginia) (Python & Tableau)
    - Contributed to developing white paper for discussion of analysis

---

PROFESSIONAL DEVELOPMENT/CONTINUING EDUCATION

- Amazon Web Service (AWS)
- Docker
- PowerBI