James Tian

jinyaoti@usc.edu | +1 213-713-0880 |https://jtian123.github.io

EDUCATION

Master of Science in **Analytics, Engineering** 08/2021-05/2023
University of Southern California Los Angeles, United States

Master of Science, **Mathematics** 02/2021-07/2021
Utrecht University Utrecht, Netherlands

Bachelor of Science, **Statistics** 09/2016-06/2020
Southeast University Nanjing, China

EXPERIENCE

Data Scientist 07/2023-Present
Keck Medicine of USC Los Angeles, United States

- Spearheaded **high-dimensional linear regression analysis** to pinpoint brain structures impacted by traffic-related air pollution, employing **causal inference** techniques to substantiate findings.
- Engineered and launched the "MRIreduce" **R package** in Github, facilitating the transformation of NIfTI format T1/FL neuroimages into structured, high-dimensional data frames, with a focus on ROI-based data pipeline processing.
- Implemented and refined the Super-Partition algorithm for robust processing of genotype data across 22 chromosomes, showcasing ability to manage and analyze **large-scale genomic datasets** effectively.
- Designed user interface with **R-Shiny** for Cisplatin Induced Hearing Loss Prediction.
- Enhanced computational efficiency by integrating **cloud computing** solutions via slurmR, significantly boosting data processing capabilities and proficiency in handling large-scale data analyses in high-performance computing environments.

Data Scientist 08/2022-12/2022
Kiana Analytics Los Angeles, United States

- Leveraged **Python** and **Spark** to streamline the processing and analysis of over 7 million geospatial data points from a UK client's factory WiFi system, enhancing data throughput and analytical precision.
- Performed **statistical modeling and analysis** to assess the occurrence frequency of MacIPAddresses within distinct spatial regions, alongside deployment of machine learning algorithms to autonomously partition individual floors into discrete sections.
- Engineered and deployed a relational database and integrated alarm system using **Python** and **SQL** for a factory setting, enabling real-time monitoring of area occupancy and secure access control.

Data Analyst 06/2022-08/2022
The American College of Financial Services Los Angeles, United States

- Utilized **SQL** to extract and process over five million rows of raw data from the company's database, supporting a comprehensive ten-year project with detailed **statistical analysis** to derive crucial business insights.
- Developed and employed **Python** to create key performance indicators, such as persistence and completion rates, and conducted **A/B testing** to explore and validate factors influencing online program completion rates.
- Designed and implemented dynamic dashboards using **Tableau**, enabling managers across multiple departments to effectively monitor and assess program procedures and outcomes, enhancing decision-making processes.

Data Scientist Co-op 02/2021-05/2021
CERN Utrecht, Netherlands

- Collaborated with Data team to clean and optimize iron particle beams data using **Pandas** and exponential rolling average method, resulting in a significant 20% performance improvement.
- Employed advanced **machine learning** technique **XGBoost** to develop powerful prediction system, achieving impressive 15% increase in prediction accuracy compared to other models.
- Leverged Variational **Autoencoder** model using **Pytorch** to address denoising challenges, achieving impressive 20% improvement in image denoising efficiency compared to other methods.

Kaggle Competition: Identifying Age-Related Conditions

- Built a cutting-edge **TabNet binary classification** model using **PyTorch** to predict medical conditions based on limited health characteristic data, enhancing existing **XGBoost** and **Random Forest** methods.

SKILLS

Python, PyTorch, R, Tensorflow, C++, Linux, SQL, Apache Cassandra, Spark, Tableau, AWS, Azure, Scipy, sklearn, RShiny, Hypothesis testing, ETL, Optimization, Machine Learning, Github