Pei Tian

(+1) 646-469-9928 | tptrix29@outlook.com | New York City, NY, 10031 | GitHub | LinkedIn | Website

EDUCATION

Columbia University Sep 2023 – May 2025 (Expected)

Master of Science in Biostatistics (Public Health Data Science) | GPA: **4.08 /** 4 *New York City, NY*

- **Core Courses:** Machine Learning (Python), Algorithms for Data Science, Data Science, Computer Systems for Data Science, Probability, Statistical Inference, Biostatistical Methods

Tongji University Sep 2019 – Jun 2023

Bachelor of Science in Bioinformatics | GPA: **4.58 /** 5, *Minor in Software Engineering* | GPA: **4.81 /** 5 *Shanghai, China*

- **Core Courses:** Data Structures (C++), Machine Learning Theory , Software Engineering, Foundation of Database, Micro-service and Web Service, Calculus, Linear Algebra, Discrete Math, Numerical Methods and Algorithms

SKILLS

Programming: **Python**, **R**, **SQL**, **Java**, C++, C#, shell, HTML/CSS/JavaScript
Data Science: **Numpy**, **Pandas**, **Scipy**, **sklearn**, **PyTorch**, TensorFlow, **tidyverse**, **ggplot2**, Shiny, Power BI
Software Development: **SpringBoot**, **MySQL**, SQLServer, MongoDB, **Docker**, React.js, Axios.js, Node.js, Bootstrap, Flask, JUnit
Concepts: **Machine Learning**, Deep Learning, **Object-Oriented Programming**, Data Structure, **RESTful API**, **RBDMS**, NoSQL, Agile Development, Cloud Computing (AWS, Google Cloud, Azure, Alibaba Cloud)

EXPERIENCE

Columbia University Irving Medical Center Sep 2023 – Dec 2023

Research Assistant *New York City, NY*

- Packed source code into an open-source **R** library while delivering correctness verification and unit testing for the algorithm blocks.
- Refactored source code with **R6 Class** to enhance the succinctness and reproductivity of source code.
- Accelerated the running time of algorithm by optimizing linear algebra calculation process and implementing core module with **C++** library like Eigen.

Shanghai Foxhub Network Technology Company August 2022 – Oct 2022

Data Engineer Intern *Shanghai, China*

- Designed relational database architecture (ER diagrams) and unstructured data source (**OSS**, Object Storage Service) on the **Alibaba Cloud** platform to support front-end webpage access.
- Managed **MySQL** database access permission and backup operation by writing **shell** scripts to maintain the stability of production and development environment.

PROJECTS

Course Management System | *SpringBoot, React.js, Bootstrap, MySQL, MongoDB, Docker, Maven, Webpack, Postman*

- Led the development of microservice-based system using **SpringBoot** and **React.js** as framework including requirement specification, system design, implementation and testing, resulting in a multi-functional and user-friendly web service application.
- Designed the database structure applying **MySQL** for relation-based data and **MongoDB** for archive-based data, while using **Docker** to maintain the isolation of different database.
- Implemented and tested RESTful API by **SpringBoot** and Postman while delivering an interactive website with **React.js, Node.js, Axios.js, Bootstrap, Webpack**.

Predicting Solid Waste Composition with Neural Network | *PyTorch, TensorFlow, sklearn*

- Implemented data cleaning, preprocessing and feature engineering for raw datasets with packages like **pandas**, **sklearn**.
- Predicted solid waste composition using neural network model while utilizing many machine learning technologies including L2 regularization, Adam optimizer and dropout layer via **PyTorch** library.
- Visualized data features and model evaluation results via **matplotlib** and **Tensorflow** library.

Neurodegenerative Diseases Onset Prediction | *Python, Flask, sklearn, MySQL*

- Completed data collection, preprocessing and feature engineering on open-source patient data about the onset of Alzheimer's disease and Parkinson's disease.
- Trained the prediction models in **sklearn** environment by classical machine learning algorithms including SVM, decision tree, etc.
- Delivered a project website with **Flask** to demonstrate details and provide an interactive prediction interface.