WENTAO LIN

New York, NY ● wentao.lin223@gmail.com ● https://wentama.github.io/ ● LinkedIn ● U.S. Citizen

EDUCATION

Case Western Reserve University   *Cleveland*, OH | Aug. 2020 - May 2024
- *Bachelor of Science in **Data Science and Analytics** (cum laude, GPA: **3.89/4.0**), Minor in **Economics***
- *Coursework:* Econometrics, Artificial Intelligence, Machine Learning for Big Data, Data Engineering, Algorithms
- *President*, Spartan Running Club; *Programmer - tama*, WRUW-FM 91.1 (radio station)

SKILLS

Programming Languages: Python, SQL, R, Java, Javascript, HTML, CSS, Stata
Tools: AWS, scikit-learn, Numpy, Pandas, Geopandas, Matplotlib, Tensorflow, Jupyter Notebooks, GEE, GIS, Ray Framework
Spoken Languages: English, Mandarin Chinese, Japanese

TECHNICAL EXPERIENCE

CTrees, Data Science and Remote Sensing Intern - *(AWS EC2, S3, Python, QGIS)*   *Pasadena, CA* | Jul. 2024 - Present
- Utilize **machine learning** to **estimate carbon globally** across forest and non-forest areas at **1-ha spatial resolution**
- **Ingest and process satellite data** for model training and deliverables, including noise **masking**, handling missing data, **harmonizing** multi-source data, and **transforming** datasets for compatibility
- **Optimize data pipeline**, reducing runtime by **98%** (from 50 hours to 1 hour) and cutting cloud computing costs by **99.6%** (from $500 to $2)

Case Western Reserve University, Undergraduate Research Assistant   *Cleveland,* OH | Jan. 2024 - May. 2024
Department of Mathematics, Applied Mathematics, and Statistics - Professor Sakshi Arya
- Developed a strategy using **reinforcement learning** to **optimize agricultural decisions** aimed at maximizing crop yield
- **Tested multiple hypotheses** to address the research problem and identified promising solutions through **data analysis**
- Implemented a **model pipeline** to integrate a crop simulation model with **R**, resulting in **improved simulation efficiency**

NASA Jet Propulsion Laboratory, Data Science Intern, UAVSAR Mission   *Pasadena, CA* | Jun. 2023 - Dec. 2023
Fall 2023 - (Python, QGIS, Unix)
- Utilized **pandas** for **data manipulation** and **analysis** to organize and merge datasets from diverse sampling techniques
- Investigated soil characteristics through **regression analysis**, including **random forest** and **support vector machines**
- Applied **ensemble learning** and **data visualization** to establish relationships between fuel moisture and polarimetric data
- **Streamlined wildfire analysis** using new data and **interpolation**, **reducing steps from 10+ to 1** and boosting efficiency
Summer 2023 - (Python, QGIS, Unix)
- **Preprocessed large-scale data** into raster format using ENVI header files and integrating various **time series datasets**
- Researched and evaluated multiple **supervised and unsupervised classification algorithms** for **image segmentation**
- Developed a **wildfire perimeter** and **burn severity classification product** with **GIS-compatible** GeoJSON outputs
- Conducted **image analysis** to extract meaningful patterns from satellite data and improve accuracy of wildfire monitoring
- Created a **GitHub tutorial**, featuring a sample fire dataset, comprehensive **documentation**, and installation instructions

PROJECTS (demo)

Timbre, a web-based music social media *(Python, PostgreSQL, Node.js, React)*   *Fall 2023*
- Designed and implemented the database, including stored procedures and indexes for **data integrity** and **query efficiency**
- Formulated **user-matching procedure** and **optimized** compatibility **score calculation** to display top matches on frontend
- Integrated **SQL and Javascript functions** to process frontend requests and **facilitate real-time database updates**

Reddit Sentiment Analysis Using Deep Learning (*Python, TypeScript, AWS*) - NLP   Spring 2023
- Designed an **API-driven data ingestion pipeline** for the full-stack application by processing 25 trending headlines from user-specified subreddits via the Reddit HTTP API, for **tokenization**, **preprocessing**, and **analysis**

ADDITIONAL EXPERIENCE

Case Western Reserve University, Teaching Assistant   *Cleveland,* OH | Sep. 2022 - May. 2024
CSDS 341 - Introduction to Database Systems (Microsoft SQL Server, PostgreSQL, My SQL) - Professor Dianne Foreback
- Utilized **hand-drawn diagrams** in office hours to effectively clarify lecture materials and homework assignments
- Designed homework problems to apply **SQL concepts**, involving drawing **ER diagrams** and **optimizing queries**