George (Gezhi) Wang

Email: gezhiwang103@gmail.com | Phone: 6072626886
Address: 222 Summerhill Dr unit 229#7, Ithaca, NY 14850

EDUCATION

CORNELL UNIVERSITY — Ithaca, New York
Master of Engineering in Systems Engineering, Computational and Data Science Track — Aug 2024 – Present
Courses: Advanced Topics in Machine Learning, Computer Vision, Database, Deep Learning, AI for Emissions, Large Scale Machine Learning System, Reinforcement Learning, Cloud Computing

DUKE UNIVERSITY & DUKE KUNSHAN UNIVERSITY DUAL DEGREE PROGRAM
Bachelor of Science in Data Science (DKU) — Kunshan, Jiangsu, China
Bachelor of Science in Interdisciplinary Studies, Data Science Track (Duke) — Durham, North Carolina
Honor: Kunshan Government Full Scholarship (merit-based), Dean's List 23 Fall — Aug 2020 – May 2024
Courses: Machine Learning, Bayesian Inference, Speech Recognition, Interdisciplinary Data Analysis, Linear Algebra, Calculus, Data Acquisition and Visualization, Probability

SKILLS

Programming Languages: Python (Pandas, NumPy, PyTorch, Matplotlib,Scikit-learn), SQL, C++.
Platform and Tools: AWS (EC2,S3), Google Cloud (Compute Engine, Cloud Storage), Docker, Git, Visual Studio Code, Tableau)

PROFESSIONAL EXPERIENCE

DATA SCIENCE INTERN (DRUG DISCOVERY) — St. Louis, Missouri
Humanwell Pharmaceutical US, Inc. — June 2025 - Present

- Developed a predictive model for Mean Squared Displacement (MSD) using SMILES molecular structures, converting SMILES into Extended Reduced Graph (ERG) fingerprints with RDKit. Implemented an ensemble learning system (SVM, Random Forest, XGBoost, MLP) achieving 60% accuracy in 5-fold cross-validation with limited data (800 samples) and deployed as a FastAPI-based inference API on the company's internal server, enabling real-time predictions via REST endpoints, saving approximately $25,000 in reagent and labor costs over six months by reducing failed experiments and manual checks.
- Designed and implemented a web-based data visualization dashboard using Python, Streamlit, Pandas, and NumPy, hosted on AWS EC2, automating experimental result recording, statistical analysis, and generating interactive real-time visualizations for researchers.
- Visualized SAR datasets using dimensionality reduction and clustering techniques with t-SNE and Linear Discriminant Analysis.

MARKET RESEARCH ANALYST EXTERNSHIP — Santa Monica, California
Mobalytics — Jun 2022 - Aug 2022

- Developed and executed SQL queries on PostgreSQL databases to extract and analyze gaming market data, building interactive dashboards in Tableau to visualize KPIs such as user growth, revenue trends, and market share.
- Created and automated data pipelines using SQL and Python (Pandas, NumPy) to consolidate competitor, partnership, and product launch data, reducing manual reporting time by 40%.
- Performed data cleaning, transformation, and statistical forecasting using Python (Scikit-learn, Matplotlib) and visualized predictive insights in Tableau, supporting data-driven strategic decisions.

PROJECTS

KAGGLE COMPETITION: CMI - Detect Behavior with Sensor Data — St. Louis, Missouri
June 2025 - Present

- Engineered a robust multi-branch model architecture, integrating 1D CNNs, GRU layers, and masked attention to process and combine features from accelerometers, gyroscopes, thermal, and ToF sensors for improved gesture classification.
- Developed and implemented advanced data augmentation strategies, including mix up, random cut and jittering, ensuring better generalization of the model across different gestures and sensor conditions.

KobeNet — Knowledge-Guided Object Bounding for Disease Detection in Chest X-rays — Ithaca, New York
Aug 2024 - Dec 2024

- Built a two-stage DL pipeline—ResNet-50 multi-label disease classifier (9 classes) trained on 43,896 images + bounding-box regressor trained on 10,755 augmented samples from 983 labeled X-rays (~1.6k boxes); engineered augmentations (scale/ flip/ translate/ brightness/ blur) to tackle severe class/box imbalance.
- Injected disease labels as priors to guide localization, improving Val IoU to 0.522 (from 0.395) and reducing Val loss to 0.0019 (from 0.0038); evaluated on NIH Chest X-ray (80/10/10 split).

MOVESAI: EMISSIONS PREDICTION BASED ON NEURAL NETWORK — Ithaca, New York
Aug 2024 - Present

- Developed a Multilayer Perceptron Model for emissions prediction, achieving an R^2 of 90%. Deployed the model using Docker for containerization and Google Cloud for scalable cloud deployment, ensuring efficient and reliable performance in production.
- Performed comprehensive data preprocessing, including handling missing values, feature selection using recursive feature elimination, log scaling emissions data, and one-hot encoding categorical variables to optimize model performance.

PUBLICATION

- **Gezhi Wang**, Zepu Wang, Peng Sun and Azzedine Boukerche, SK-SVR-CNN: A Hybrid Approach for Traffic Flow Prediction with Signature PDE Kernel and Convolutional Neural Networks . In *IEEE International Conference on Communications*, 2024