Haozhen Shen

 LinkedIn | 📱 647-505-0656 | 🌐 haozhenshen.com | ✉ haozhen.shen@outlook.com | ⭕ GitHub

Skills

- Python | C | C++ | Java | JavaScript | TypeScript | Node | Bootstrap | Express | Next | React | Postman | MongoDB | PostgreSQL
- Unit Testing | OOP | Kafka | Redis | Distributed Systems | Kubernetes | Docker | CI/CD | Jenkins | Git | Bitbucket | Jira | AWS
- R | NumPy | Pandas | scikit-learn | Prophet | XGBoost | CatBoost | LightGBM | Pytorch | Pytorch Lightning | Tensorflow | Keras
- Hugging Face | Transformers | Generative Modeling | VAE | EBM | W&B | Neptune | MLFlow| ONNX | Kedro | Airflow | Spark | Ray

Experience

Machine Learning Scientist — **BluWave-ai** — *Ottawa, ON, Canada* — **05/2022 - Current**

- Lead Scientist for energy forecasting products.
- Built an end-to-end machine learning pipeline for time series forecasting of electricity load data using **ECS, Elasticache, Lambda, Ray, Kedro, Airflow, MLFlow, Tensorflow, Pytorch,** and **LightGBM** resulted in a 22% improvement in performance (MAE) and an 80% reduction in inference time compared to the previous method.
- Led the development of the ML system for load prediction products. From identifying system requirements and partner dependencies, and the entire life cycle of data engineering, model development, testing, and configuring metrics, alarms, and dashboards (**Grafana**).
- Took the initiative to refactor existing machine learning pipelines into modularized components leveraging Kedro and introducing MLFlow to the team accelerating model development and lowering the cost of maintenance.

Research Assistant — **University of Toronto** — *Toronto, ON, Canada* — **02/2022 - 06/2022**

- Applied deep learning methods to principal-agent mean-field games. Modeling the Renewable Energy Credit market using McKean-Vlasov Forward-Backward Stochastic Differential Equations. Solving the stochastic control problem using deep learning.

Data Mining Engineer, Intern — **CRRC Academy** — *Beijing, China* — **04/2019 - 08/2019**

- CRRC is the world's largest rolling stock manufacturing company.
- Joined the algorithm team dedicated to analyzing the vehicle's operation condition for rail networks.
- Accelerated data preprocessing pipelines using MATLAB which increased preprocessing speed by 33%.
- Designed an LSTM-CNN classification algorithm to validate and identify potential vehicle failures.
- Developed a threshold analyzing algorithm, which helps distinguish valid data from noise caused by a sensor failure.

Projects

YelpCamp (Full-Stack, Web Development)

- A campsite review and rating website for travelers, implemented both front and back end and deployed on Heroku. (JavaScript, Node, Express, React, MongoDB). https://github.com/haozhenshen/YelpCamp.

Adaptive Noise Score Network (Machine Learning, Generative Modeling)

- Designed a Score Based Generative Model, inspired by Adaptive MCMC techniques. Achieving competitive results in image generation.

- **Mini Redis**: Build a simplified version of Redis in C++ that handles multiple concurrent clients with Echo, Set, and Get commands.
- **Feedback Prize - NLP**: Fine-tuning Deberta models to assess the language proficiency of 8th-12th grade English Language Learners.
- **Energy-Based VAE**: Image generation by jointly training VAEs and Energy Based models (EBMs) through Contrastive divergence.
- **File Systems**: Build an ext2 file system in C including disk formatting operations.

Education

Master of Science — **University of Toronto (St. George)** — *Toronto, ON, Canada* — **09/2021 - 04/2022**

- Statistics, Focus on Generative Modeling, Probabilistic Models, and Statistical Learning Theory.

Bachelor of Science — **University of Toronto (St. George)** — *Toronto, ON, Canada* — **09/2016 - 04/2021**

- Double Specialist (Major + Focus), Computer Science, and Applied Mathematics.

Mentorship

- UTFUN Tutor: Tutored CSC108 to undergraduate students responsible for teaching basic algorithms and data structures in Python.

Others

- **National second-level athlete (**three-dan**)** in the game of GO.