

Haozhen Shen

[in LinkedIn](#) | [647-505-0656](tel:647-505-0656) | haozhenshen.com | haozhen.shen@gmail.com | [GitHub](#)

Skills

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

Experience

Machine Learning Scientist

BluWave-ai

Ottawa, ON, Canada

05/2022 - 06/2023

- Exceptionally orchestrating the full forecasting product life cycle. Starting from managing the extraction and transformation of data from a myriad of sources.
- Built an end-to-end machine learning pipeline for time series forecasting of electricity load data using **Kinesis Data Streams, Lambda, Ray, Kedro, Airflow, MLFlow, Tensorflow, Pytorch**, and **LightGBM** resulting in a 22% improvement in performance (MAE) and an 80% reduction in inference time compared to the previous method.
- Implemented rigorous testing protocols to ensure a 90% reduction in errors. Configured key metrics and alarms for precise monitoring, craft insightful and interactive dashboards using Grafana, and ensure seamless deployment.
- Improved the backtesting infrastructure for time series data with a 33% increase in efficiency, accelerating model development and testing.
- Took the initiative to refactor existing machine learning pipelines into modularized components leveraging Kedro accelerating model development and lowering the cost of maintenance.

Research Assistant

University of Toronto

Toronto, ON, Canada

02/2022 - 06/2022

- Applied advanced deep learning techniques to design and implement function approximators, effectively solving complex stochastic control challenges. These solutions were successfully deployed to tackle intricate Renewable Energy Credit market models within the principal-agent mean-field game framework. This approach yielded a 10% increase in model accuracy, optimizing decision-making and enhancing strategic outcomes.

Data Mining Engineer, Intern

CRRC Academy

Beijing, China

04/2019 - 08/2019

- Contributed expertise to the algorithm team at CRRC, a global leader in rolling stock manufacturing.
- Enhanced the data preprocessing pipelines by leveraging MATLAB and Python, resulting in a remarkable 33% acceleration in data processing speed.
- Developed a robust end-to-end data pipeline for a new algorithm developed for validation and early identification of potential vehicle failures, directly improving operational reliability.
- Implemented a threshold analyzing algorithm in effectively discerning valid data from sensor-related noise stemming from potential sensor failures. This solution streamlined data interpretation, contributing to a 33% reduction in false alarms and bolstering overall system efficiency.

Software Engineer, Intern

Shanda Interactive Entertainment

Shenzhen, China

04/2018 - 08/2018

- Collaborated seamlessly with the product management team and developers to architect and execute a robust application monitoring system, capturing critical player data to drive informed decisions.
- Implemented an internal A/B testing framework, enhancing experimentation and data-driven optimization.
- Addressed various bugs on existing websites and applications in production that have been present for years.
- Optimized alerting systems for swift response to player-reported bugs, reducing response times by 50%.

Education

Master of Science

University of Toronto

Toronto, ON, Canada

09/2021 - 04/2022

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

Bachelor of Science

University of Toronto

Toronto, ON, Canada

09/2016 - 04/2021

- Double Specialist, Computer Science, Applied Mathematics.

Certifications

- AWS Certified Solutions Architect - Associate

Projects

- **MarketSentinel:** A SAS for stock market sentiment visualization. A DistilRoberta model from **Huggingface** and fine-tuned on financial news is used to perform text sentiment classification for news data leveraging the **Huggingface** Inference API. [GitHub Link](#)