**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 | W&B | ONNX | Kedro | Airflow | Spark | Ray

• **Platform:** AWS | Databricks | Vercel

**Experience** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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

• 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.

• Involved in the entire life cycle of forecasting products. From **ETL** data from various sources, conducting statistical analysis on data, model development, testing, configuring metrics, alarms, dashboards (**Grafana**), and deployment.

• Improved the backtesting infrastructure for time series forecasting projects, 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** |

• Implemented function approximators to solve stochastic control problems using deep learning. The control problem models the Renewable Energy Credit market in the principal-agent mean-field game setting.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data 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 and Python which increased preprocessing speed by 33%.  
• Implemented an end-to-end data pipeline for 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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Software Engineer, Intern** |  | **Shanda Interactive Entertainment** | *Shenzhen, China* | **04/2018 - 08/2018** |

• Worked with the product management team and developers to build application monitoring player data.

• Implemented an internal A/B testing framework in addition to the player data monitoring system.

• Addressed various bugs on existing websites and applications in production that have been present for years.

• Optimized alerting systems when receiving a high volume of player-reported bugs.

**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** |

• Specialist, Computer Science, Machine Learning path.

• Specialist, Applied Mathematics, Probability and Statistics path.

**Projects** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**MarketSentinel (Full Stack Machine Learning, NLP)**

• A stock market sentiment visualization website. 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. A black cat with a blue circle

Description automatically generated with low confidence [GitHub](https://github.com/haozhenshen/marketview) Link

• **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.

• **Adaptive Noise Score Network**: Designed a Score Based Generative Model, inspired by Adaptive MCMC techniques.

• **Energy-Based VAE**: Image generation by jointly training VAEs and Energy Based models (EBMs) through Contrastive divergence.

**Others** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
• **National second-level athlete** in the game of GO.

• **First Place**: Ranked first place in the three-dan promotion competition of the game of GO in, Shenzhen, China.