David Lu

david.linyi.lu@gmail.com • david5010.github.io • 650-305-6097

EXPERIENCE

Skylar Capital Management LP Houston, TX
Data Scientist for Natural Gas Market *June 2024 - Present*

- Engineered Airflow pipelines to automate and orchestrate machine learning weather forecasting models, reducing data acquisition time by several hours and providing earlier insights than traditional methods
- Conducted extensive research to repurpose and enhance machine learning weather forecasting models, implementing error correction techniques to improve prediction accuracy and reliability
- Developed and implemented renewable energy forecasting models for ERCOT's wind and solar outputs, enhancing the accuracy of energy supply predictions to inform natural gas trading strategies

Baldi Lab Irvine, CA
Machine Learning Researcher *February 2023 - June 2024*

- Designed various architectures such as DeepSet, and Set Transformer that can predict the quality of an antenna array
- Improved existing antenna pattern design by 89% by creating a custom gradient descent in PyTorch
- Boosted pattern cost calculation efficiency by 5000% by transitioning MATLAB code to PyTorch Tensors

EDF Innovation Lab Palo Alto, CA
Machine Learning Engineer for Energy Markets - PyTorch, Pandas, Airflow, Xarray *September 2022 - January 2024*

- Implemented various deep-learning models with time dependencies to forecast electricity load based on climate data
- Explored Dask and SageMaker integration for enhanced data processing and efficient model training with large datasets
- Developed an algorithm that integrates population density data to enhance temperature-dependent load estimation
- Expedited the team's research processes by over 30%, by engineering spatial data pipelines using Xarray and GeoPandas
- Predicted natural gas outages by analyzing the relationship with extreme temperature and spatial relations using Xarray

Lyrid Irvine, CA
Software and Data Engineer Intern - Airflow, PyMongo, Spark *December 2021 - September 2022*

- Designed an internal API to easily retrieve user data from MongoDB and facilitate data dumping for user activity analysis
- Improved the features of an authentication service by utilizing graphene and Django to create GraphQL mutations
- Saved the team 20 hours by resolving two critical bugs for Sheliak by implementing test cases using Insomnia
- Reduced development time by 8% by prototyping a scalable ETL pipeline using Airflow, Spark and Delta Lake

PROJECTS

Efficient Generative Models PyTorch/Triton/CUDA
Exploring pre-training pruning methods and performance gain on generative models

- Implemented the Gradient Signal Preservation (GraSP) technique from a machine learning paper to prune neural networks pre-training, adapting the methodology to generative models using PyTorch
- Explored advanced GPU programming methods using the Triton language, optimizing neural network operations to achieve substantial performance gains in model training and inference times

SKILLS & INTERESTS

Language: Fluent in English, French, Mandarin

Languages/Framework: Python, Julia, R, SQL, C/C++, MATLAB, PyTorch, Scikit-learn, Tensorflow, Numpy, Pandas, Apache Spark, Dask, Airflow, Xarray, GeoPandas

Software: Linux, AWS, Docker, MongoDB, PostgreSQL, Git

Interests: Finance, Climate Science, Computer Vision, NLP, Art, Sailing

EDUCATION

University of California, Irvine | GPA: 3.86/4.00 Irvine, CA
B.S. Computer Science, Statistics; Specialization: Machine Learning and Statistical Methods
Relevant Courses: Data Structures, Machine Learning, Graphical Models, Generative Models, Bayesian Inference
Extracurriculars: Vice President of Technology - **Alpha Kappa Psi** | Education Director - **Commit the Change (CTC)**