David Lu

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

EXPERIENCE

EDF Innovation Lab Palo Alto, CA
Data Scientist for Energy Markets - PyTorch, Pandas, Airflow, Xarray *September 2022 - Present*

- 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

Yes! Star Corporation Shanghai, SH
Data Science Intern - Pandas, Numpy, Seaborn *June 2019 - July 2020*

- Reduced the data noise of the company's projected sales regression model by cleaning 36 datasets using Pandas
- Saved 50 hours for the marketing team by automating the visualization of hundreds of datasets using Seaborn
- Increased accuracy of market analysis by 10% by finding target market data by web scrapping various websites

RESEARCH

University of California, Irvine - Deep Data Lab Irvine, CA
Normalized Metered Energy Consumption (NMEC) - Advisor: Prof. Matthew C. Harding *September 2023 - Present*

- Conducted NMEC research using Pecan Street with NOAA temperature statistics for enhanced energy usage prediction
- Developed PyTorch models to assess energy savings in programs through counterfactual scenario analysis
- Employed OpenEEmeter for benchmarking and validation of energy savings calculated by the PyTorch model, ensuring compliance with industry-standard efficiency measurement methods and enhancing the reliability of the results

University of California, Irvine - Baldi Lab Irvine, CA
Antenna Array Optimization - Advisor: Prof. Pierre Baldi *February 2023 - Present*

- Designed various architectures such as DeepSet, and Set Transformer that can predict the quality of an antenna array
- Implemented the preprocessing stage of the project by creating custom datasets in PyTorch for 1D and 2D representation
- Created a custom gradient descent function that optimizes the inputs, giving an antenna array pattern of higher quality
- Boosted pattern cost calculation efficiency by 5000% by transitioning MATLAB code to PyTorch Tensors

EDUCATION

University of California, Irvine Irvine, CA
B.S. Computer Science, Statistics; GPA: 3.85/4.00 *Graduating June 2024*

CAMPUS INVOLVEMENTS

Alpha Kappa Psi
Technology Chairman *October 2020 - Present*

- Automated application processing by using App Script, saving 40 hours per quarter and reducing errors by 90%
- Taught 17 members data analysis tools such as Pandas, Matplotlib and Numpy by organizing weekly workshops

SKILLS

Programming Languages: Python, Julia, R, SQL

Framework: PyTorch, Scikit-learn, Tensorflow, Numpy, Pandas, Apache Spark, Dask, Airflow, Xarray

ADDITIONAL

Languages: Fluent in English, French, Mandarin

Interests/Hobbies: Sailing, cooking, art, horse betting, horology, Machine Learning