Frank Kloster

 frankkloster | frank-kloster-a8940a149 | frankkloster.github.io
 (805) 345-6421 | frankkloster@gmail.com

SKILLS

Programming : Python, R, C/C++, SQL, bash. Data visualization with ggplot2 and Matplotlib. Machine learning and statistics with Scikit-Learn, Statsmodels, Numpy, Pandas, PyTorch, Pytorch Geometric, and HuggingFace. Web design with Flask. Documentation generation with Sphinx.

Tools: bash, GNU core utils suite, VSCode, git, ssh, vim, sed, grep, awk.

Machine Learning : KMeans, Gaussian Mixture Models, PCA, Autoencoders, DBSCAN, UMAP, TSNE, Logistic Regression, XGBoost, Random Forests, Support Vector Machine.

Deep Learning : Self-supervised Learning, Contrastive Learning, Transformer Architectures, RNNs, CNNs

LLMs : Retrieval augmented generation (RAG), prompt engineering, agent based models, knowledge distillation.

Statistics : Bayesian and Frequentist Inference, Markov Chain Monte Carlo Methods, Hypothesis Testing, and Exploratory Data Analysis.

EXPERIENCE

Hitachi America   Aug. 2022 – Apr. 2024
Machine Learning Research Scientist   Santa Clara, CA

- Engaged in research at Stanford University, co-developing a time-dependent deep graph convolutional neural network to predict supply chain transactions.
- Designed and implemented an agent-based large language model (LLM) tool, automatically translating natural language into machine code, empowering factory workers with no coding experience to generate visualizations of critical factory metrics.
- Developed a self-supervised learning technique for video analysis, enabling researchers to leverage unlabeled data. This streamlined the labeling process, significantly improving research efficiency.
- Leveraged Flask to rapidly develop Proof-of-Concepts (PoCs) for various business unit projects. Utilized serverless deployments on AWS Lambda to facilitate quick demonstrations and effective communication of project ideas.

Transamerica   Dec. 2019 – Mar. 2021
Data Scientist   Denver, CO

- Led efforts to automatically generate leads to life insurance agents based on geospatial and other parameters.
- Worked extensively to maintain internal packages used by a wide variety of data science teams across Transamerica to ease development.
- Evaluated multiple data sets in order to aid underwriting and actuarial efforts to better understand mortality risks.
- Used programming languages such as Python, R, and C++. Used Python libraries such as PyTorch, Scikit-Learn, Pandas, Matplotlib, Survival
- Used technologies such as Jupyter Notebooks, RMarkdown, AWS, Domino, Sphinx, and Jenkins.

Hitachi America   Jun. 2017 – Sep. 2017
Research Intern   Santa Clara, CA

EDUCATION

The Data Incubator   2019
Data Science Fellow   Online

University of California, Riverside   2019
PhD Mathematics   Riverside, CA

University of California, Santa Barbara   2011
B.S. Mathematics and Physics   Santa Barbara, CA