Xiangyi Kong, Sam

Berkeley, CA | (925) 212-9676 | kxy118@gmail.com | [LinkedIn](#) | [GitHub](#)

EDUCATION

University of California, Berkeley Berkeley, CA

Bachelor of Art in Data Science 2021/08 - 2022/12

Relevant Coursework: Probability, Statistical Inference, Machine Learning, Data Mining

Diablo Valley College Pleasant Hill, CA

Associate in Science in Computer Science 2019/01 - 2021/05

Relevant Coursework: Linear Algebra, Python, C & C++, Differential Equations, Discrete Math

WORK EXPERIENCE

EchoPlus AI LLC, *Machine Learning Engineer*, Boulder, CO 2024/06 - 2024/12
- Conducted features by extracting user behavior, contextual signals, and textual attributes, applying **Word2Vec** embeddings, **TF-IDF** dimensionality reduction, and cross-feature combinations to enhance representation.
- Developed and optimized a multiple-recall strategy using item-based collaborative filtering, user-based CF, embedding similarity, cold-start solutions, and YouTubeDNN models, achieving enhanced recall through weighted averaging and FAISS-accelerated inference.
- Built ranking and classification pipelines with **LightGBM**, achieving state-of-the-art **NDCG of 0.98** via cross-validation, improving AUC by 10% through Deep Interest Network (DIN) and stacking ensemble methods.
- Designed scalable, high-performance recommendation workflows using **TensorFlow** and **DeepCTR**, handling diverse user preferences and optimizing computational efficiency.

Dalian United International Shipping Agency Co., Ltd., *Data Scientist*, Remote 2023/06 - 2023/12
- Conducted **sentiment analysis** on reviews using **TextBlob** to extract polarity and subjectivity scores. Enhanced feature engineering with part-of-speech tags, named entities, term frequency, and text complexity via **Textatistic**, achieving a **17% improvement in rating precision**.
- Built a user network graph with **NetworkX** from merchant data, identifying **4 influential communities** in Stuttgart using community detection algorithms and visualized relationships with spring layout graphs to reveal user behavior.
- Integrated **large language models** (GPT-3) for advanced text analysis, enabling deeper sentiment interpretation and summarization of reviews, reducing manual analysis time by 20%.

PROJECT EXPERIENCE

Stock Market Customer Classification 2024/01 – 2024/03
- Designed and implemented an RFM (Recency, Frequency, Monetary) analysis pipeline in Python to segment e-commerce customers into 8 distinct tiers, applying optimized vectorized operations for scalability on large datasets.
- Implemented personalized marketing campaigns for each segment increasing 5% customer retention rates.
- Developed and deployed K-Means clustering models to identify customer patterns, using the Elbow Method and Silhouette Score to fine-tune the optimal number of clusters, improving segmentation accuracy by 15%.

Financial Product Capital Flow Forecast 2023/09 – 2023/12
- Extracted time-series features (e.g., time differences and sliding window features) and employed **SHAP values** and **permutation importance** to identify crucial predictors, reducing model complexity and achieving a 9% improvement in Absolute Error (AE).
- Experimented with a comprehensive suite of regression models (Linear Regression, Random Forest, XGBoost, SVR), employing grid search to fine-tune hyperparameters and selecting Random Forest for its superior performance.
- Leveraged **Prophet** for 30-day cash flow predictions, incorporating scenarios with and without holidays to capture seasonal trends, and visualized results using Matplotlib to communicate findings effectively.

SKILLS AND INTERESTS

- *Programming Skills:* Python(Pandas, Numpy, Tensorflow, PyTorch, PySpark, Sklearn), R, Oracle, C++, Java, SAP, Web Technologies (HTML, CSS), MS Access, MS Excel, Power BI, Tableau, SQL, ETL, AWS
- *Analytics Skills:* Stats (Hypothesis Test, A/B Test, Regression), Traditional Machine Learning (Supervised, Unsupervised, Ensemble Methods), Deep Learning (CNN, RNN), NLP (Sentiment Analysis), LLMs (Prompt Engineering), Recommendation System (Collaborative Filtering, Content-Based Filtering)
- *Interests:* Swimming, Basketball, Soccer, Golf, Lego, Bartender, Poker