Qiaoling Ye

Qiaoling (pronounced "chow-ling") (310) 721-5381 yeqiaoling170@gmail.com linkedin.com/in/qlye github.com/yeqiaoling

SUMMARY

Machine learning engineer with a Ph.D. in Statistics and 4 years of industry experience. Specializing in building scalable machine learning pipelines, recommendation systems, and fine-tuning large language models (LLMs). Proficient in deploying ML models using MLOps best practices.
Programming: Scala, Spark, Python, PyTorch, TensorFlow, Scikit-Learn, Spark MLlib, FastAPI, airflow, Jenkis, Docker

EMPLOYMENT

Supplyframe / Siemens - *Data Scientist* *Nov. 2021 – Present | Pasadena, CA*

E-commerce Machine Learning

- **Personalized Recommendation System:** Developed a GNN-based ranking model in *PyTorch*, optimizing cold-start engagement, tuning hyperparameters, and boosting CTR by **3X** ($0.3\% \mapsto 0.9\%$). Provided ranked lists for A/B testing, leading to a new recommendation section launch.
- **Bot Detection & Fraud Prevention:** Developed and deployed an ensemble-based bot detection method combining supervised and unsupervised learning, achieving a **>95% AUC score**. Automated bot candidate generation using a feature engineering pipeline, reducing manual review workload by **41%**.
- **User Segmentation & Targeting:** Built a data pipeline for eCommerce user segmentation using *Scala Spark*, integrating it with marketing APIs to enable real-time targeted campaigns.

LLM & AI Model Optimization

- **LLM-powered Data Standardization for Electronics Parameters:** Fine-tuned *GPT-4o-mini* on **300 annotated samples** to normalize unstructured part descriptions. Achieved **94% token accuracy** and removed **40% redundant fields** via consistent taxonomy mapping. Deployed via *FastAPI* with `/batch_standardize/` and `/health_check/` endpoints to support scalable data integration.
- **LLM Model Evaluation:** Built an embedding-based evaluation pipeline (*BERT*, *Summary Consistency*) to benchmark LLM faithfulness, detecting **2%** of records with high hallucination probabilities.

Monitoring & Infrastructure

- **Traffic Anomaly Detection System:** Designed and implemented an *Airflow*-scheduled ML pipeline to detect significant traffic anomalies across country-product combinations in real time. Integrated with alerting systems, achieving **98%** outage detection accuracy and reducing system downtime by **66%** (from 3 days to at most 1 day).
- **Automated Jobs Scheduling:** Designed and scheduled *Pricing and Recommendation API* automated jobs to run every other Monday on *Jenkis*, ensuring timely data updates and reducing manual operations.
- **ML Model Deployment:** Developed scalable ML pipelines using *Spark MLlib* and stored trained models in *HDFS* for large-scale inference. Transitioned from **monthly batch processing (30-day delay)** to **event-driven updates (2-hour latency)**, achieving over **350**× improvement in data freshness and enabling near-real-time analytics.

Amazon - *Data Scientist Intern* *Jun. 2019 – Sep. 2019 | Seattle, WA*

Supply Chain Optimization

- **Automated Inventory Evaluation:** Developed an ML-based automation pipeline to analyze inventory performance, identifying (up to) top 5 profit-driving attributes and reducing manual report generation time by **2 hours per week**.

Universal Music Group - Data Scientist Intern (Part-time) *Feb. 2019 – Jun. 2019 | Santa Monica, CA*

Time Series Forecasting

- **Demand Forecasting:** Delivered investment recommendations to marketing teams by analyzing stream demand with advanced *time series models*, including Exponential Smoothing, ARIMA, Prophet, and Bayesian Structural Time Series.

PUBLICATIONS

Qiaoling Ye, Amini A. Arash, and Qing Zhou. Federated Learning of Generalized Linear Causal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 10, pp. 6623–6636, Oct. 2024.

Qiaoling Ye, Amini A. Arash, and Qing Zhou. Optimizing Regularized Cholesky Score for Order-based Learning of Bayesian Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, pp. 3555–3572, Oct. 2021.

EDUCATION

University of California, Los Angeles (UCLA)

Doctor of Philosophy (Ph.D.) in Statistics, GPA: 3.95/4.00, Thesis: Order-based Learning of Bayesian Networks *2015 – 2021*

Bachelor Science (B.S.) in Financial Actuarial Mathematics, GPA: 4.00/4.00 *2012 – 2014*