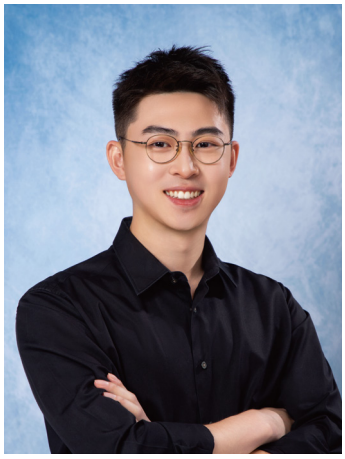

Ziqi Xiao 肖子奇

Phone/Wechat：18510186056
xzqshawn@126.com
GitHub: https://github.com/ZiqiXiao

I am a Python developer with a profound grasp of architectural design concepts, exhibiting a clean code style. I bring a substantial wealth of experience in full-stack development, Proficient with backend frameworks such as Flask, Django, and Fastapi. At the same time, I specialize in machine learning and deep learning realms. My adaptability, rapid learning capability and problem-solving skills set me apart.

Education Background

The University of Iowa

Majors: Mathematics, Actuarial Science
Relevant Courses: Advanced Mathematics, Probability Theory and Mathematical Statistics, Financial Economic Mathematical Modeling, Time Series Analysis, etc.
Degree: Double Bachelor's Degree
GPA: 3.8/4.0

Technical Stack

Machine Learning: Sklearn, LightGBM, XGBoost, LinearRegression, RandomForest, etc.
Deep Learning: PyTorch, TensorFlow; including CNN, DNN, Twin Towers, Wide&Deep, etc.
Full-stack Development: Proficient in Python, Flask, Django, Fastapi. Familiar with Java, JavaScript, TypeScript, SpringBoot, Vue. Expert in SQL, Redis, MySQL, Hive. Knowledgeable in NOSQL solutions such as MongoDB, ES, and Pinecone vector databases.
Cloud Services: Terraform, AWS, DigitalOcean, Vultr.
DevOps: Git, Gitlab, Linux, CI/CD, Docker.

Project

Craiditx Anti-fraud Product Development & Deployment (Data Scientist)

Sep 2023 - Present

This product leverages user transaction behavior and machine learning to provide banks with warnings for blacklisted accounts and risk account identification.

- Spearheaded the code recovery and test framework establishment, creating a pytest-based SQL code testing process. This reduced post-recovery code testing times, establishing a standardized workflow. Employing object-oriented programming and dependency injection, I crafted a comprehensive test package.
- Collaborated in the research, development, and optimization of the blacklist identification device, enhancing LightGBM's sorting accuracy and incorporating logging and multi-threading functionalities to the SQL batch processing module.

Machine Learning Platform for a Military Hospital in Kunming (Full-stack Developer + Machine Learning)

Apr 2023 - Aug 2023

A machine learning platform designed for drug dosage studies, allowing training, tuning, deployment, and application of various machine learning and DNN deep learning models through a graphical interface.

- Single-handedly managed the project: used Vue (Element admin) for the frontend, Springboot for backend model and user data management, MySQL for database storage, Flask for machine learning API service, integrated KeyCloak for user permissions, and utilized Docker for service deployment.
- Provided five fundamental models and a PyTorch-based DNN model, alongside SpuerLearner. Features included real-time model training visualization, performance metrics, and basic parameter tuning.
- Implemented data template reuse, model management (storage to deployment), logging, and user management functionalities.

Content Recommendation Project for China Merchants Bank's App (Data Scientist)

Sep 2021 - Feb 2023

Developed a recommendation system for financial products and content within the app, optimizing the system iteratively for various business modules.

- Consumer-Content User Profile Tags: Mapped user credit card consumption data to content interest, creating multi-level interest tags. Also, combined user app activity data to predict user interests using the LightGBM algorithm, initializing T+1 content push management.
- Other Modules: Article recommendation (word vector-based), knowledge graph recommendation system (neo4j).

Octopus Recommendation Central Platform for SPD Bank (Data Scientist)

Jul 2021 - Sep 2021

Enhanced the bank's financial product recommendation system by refining the algorithms, thereby improving recommendation accuracy and diversity.

- Managed system design, algorithm optimization, data exploration, strategy design, code development, and report compilation.
- Introduced multi-method recall at the recall layer: label recall, vector similarity recall, associated event recall, popular recommendations, and user profile recalls. Used HiveSQL for data cleaning and structuring.
- Employed XGBoost for preliminary sorting of the entire user recall pool, retaining the top 100 results. For high-value users, a Deep&Wide model in PyTorch was used for fine-tuning, retaining the top 20.
- The model increased the top 5 conversion rate by 15 times compared to organic traffic and enhanced product diversity by 5 times.

Other Projects: Private banking product recommendation for SPD Bank, loan product recommendation for Zhejiang Rural Credit Cooperatives, loan recommendation for SPD Bank, product deployment for Bank of China's Shenyang branch, product deployment for Postal Savings Bank of China's Shanghai branch.

Professional Experience

Shanghai Craiditx Information Technology Co., Ltd (Data Scientist)

Apr 2021 - Present

- Led teams to execute projects for major banks in China, including Bank of China, China Merchants Bank, Postal Savings Bank, and SPD Bank. Projects encompassed product/content recommendation, anti-fraud measures, and risk control. Responsible for technological road mapping, requirements alignment, project management, and results presentation. Actively participated in project implementation, collaborating with team members. Accumulated project values reached 7 million RMB.
- Leveraged Python and SQL for tasks including data cleaning, feature engineering, model construction, training, testing, tuning, prediction. Utilized Linux servers or client platforms for code deployment and live application, addressing various technical challenges along the way.
- Developed the anti-fraud product, overseeing code recovery, testing framework, and batch scheduling content research. Enhanced product workflows, saving 100% of deployment time and 70% of code testing recovery time on average.

Fresh Fresh Trading (Shanghai) Co., Ltd (Data Analyst)

Oct 2020 - Mar 2021

- Used SQL for daily operational data extraction, continuously monitoring projects and suggesting improvements. Utilized Python libraries like Pandas, numpy, and seaborn for data cleaning, aggregation, and visualization. Deployed RandomForest from sklearn for churn prediction.