Qi Xiang

+44 (0) 7857095426 │ qixiang.aubury@gmail.com │ LinkedIn │ Personal Website

EDUCATION

Imperial College London *Sep 2024 - Present*
- MSc Environmental Data Science and Machine Learning
- Relevant Courses: Deep Learning, Big Data Analytics, Data Science and Machine Learning

The University of Manchester *Sep 2021 - Jul 2024*
- BSc Computer Science - First Class Honours
- Relevant Courses: Data Science, Machine Learning, Natural Language Processing, Computer Vision

INTERNSHIP EXPERIENCE

Zilliz *August 2024 - September 2024*
Marketing Intern (Research Focus)

- Researched vector indexing algorithms such as HNSW and IVF for high-dimensional vector search, analyzing trade-offs in performance and resource usage.

- Explored RAG (Retrieval-Augmented Generation) system architecture and integration with the LangChain framework; proposed optimization strategies for vector search to support product development.

- Collaborated with R&D and marketing teams to conduct requirements analysis and technical feasibility studies, mapping product features to market needs.

MCM China *July 2023 - August 2023*
Information Technology Intern

- Developed an automated financial reporting system integrated with Power BI for enhanced efficiency in financial data analysis.

- Designed and implemented a modular ETL pipeline to automate data extraction, cleaning, transformation, and loading, ensuring stable and efficient data processing.

- Created multi-dimensional visualizations in Power BI for trend analysis, variance comparison, and forecasting, supporting strategic decision-making for management.

RESEARCH EXPERIENCE

Research Project: Learnable Agent for Video Game Playing (Deep Reinforcement Learning)
August 2023 - May 2024 GitHub: https://github.com/Aubur9y/Learnable-Agent-for-Sekiro

- Developed an autonomous agent with PyTorch capable of mastering video games using deep reinforcement learning.

- Researched, implemented, and optimized cutting-edge algorithms, including Deep Q-Networks (DQN), Dueling DQN, and Proximal Policy Optimization (PPO), optimizing performance by incorporating an entropy term into PPO's loss function to enhance exploration and mitigate premature convergence.

- Integrated Convolutional Neural Networks (CNNs) with custom architecture layers, leveraging frame-skipping techniques to minimize data redundancy, accelerate training processes, and facilitate real-time strategic decision-making, significantly improving gameplay performance.

- Architected a customized testbed environment from scratch, enabling comprehensive benchmarking and rigorous evaluation of reinforcement learning algorithms for continuous performance optimization.

- Engineered a robust reward mechanism, incorporating techniques like reward shaping, discount factors, and entropy-based loss functions, increasing learning efficiency by 45% and delivering more robust, adaptive gameplay strategies.

PROJECT EXPERIENCE

Customer Segmentation & Recommendation System
Github: https://github.com/Aubur9y/Recommendation-System-based-on-Customer-Segmentation

- Cleaned raw transaction data by handling missing values, removing duplicates, and detecting outliers. Engineered key features based on RFM analysis (Recency, Frequency, Monetary), along with shopping habits, cancellation rates, seasonality, and spending trends.

- Applied Isolation Forest to detect anomalous users; performed data standardization and Principal Component Analysis (PCA) to reduce dimensionality and address multicollinearity while preserving key information.

- Performed K-Means clustering on PCA-transformed data. Determined the optimal number of clusters using Elbow and Silhouette methods, and evaluated clustering performance using Silhouette Score, Calinski-Harabasz Index, and Davies-Bouldin Index with support from 3D visualizations and radar charts.

- Developed a personalized recommendation system by identifying unpurchased popular items within each user segment, aiming to increase conversion rates and reduce order cancellation.

Personalized Product Recommendation System
GitHub: https://github.com/Aubur9y/Personalised-Product-Recommendation-System

- Integrated raw product data and performed feature engineering to construct key attributes (e.g., discounted price, ratings, user engagement, product categories), effectively capturing product characteristics and user preferences while addressing data inconsistency and redundancy.

- Built a content-based filtering model using cosine similarity and nearest-neighbor search on standardized product features, enabling accurate item matching and overcoming limitations of single-method recommendations.

- Developed a collaborative filtering model based on user-item interaction data, using efficient indexing to identify similar users and mitigate data sparsity and cold-start issues. Combined dynamic weighting and category diversity penalties to create a personalized hybrid recommendation system, aimed at significantly improving user conversion rates and overall platform engagement.

Flood Risk Prediction

- Conducted exploratory data analysis (EDA) and designed two key data processing pipelines to ensure efficient and reliable data preprocessing.

- Developed and trained multiple machine learning models, including Random Forest Classifier and XGBoost; conducted performance comparisons and utilized RandomizedSearchCV for hyperparameter tuning to improve model accuracy and generalization.

- Led code integration efforts, consolidating independently developed modules from an 8-member team into a clean, modular, and well-structured codebase that adhered to project standards and enhanced maintainability.

SKILLS

- **Programming Languages & Tools:** Python, Java, SQL
- **Machine Learning & Data Science:** PyTorch, Scikit-learn, Pandas, NumPy, Transformers, DIN, CNN, LightGBM, LR, FM, GBDT, DQN, PPO
- **Data Visualization & Business Intelligence:** Power BI, Matplotlib, Seaborn
- **DevOps & Cloud:** Docker, Git