XIAOSHI LU (ALEXANDRA)

Phone: (+61)420970586 | Email: sashkalu1777@gmail.com | LinkedIn: LinkedIn Xiaoshi Lu | Location: Sydney, NSW, Australia

As a data scientist with 5 years of experience, I have developed strong professional skills, including proficiency in Python and SQL programming, statistical analysis, machine learning, and Generative AI, as well as effective stakeholder management in various complex commercial environments. I am now seeking opportunities to contribute to impactful projects and grow into a leadership role.

SKILLS

- **Programming & EDA:** Python (Pandas, NumPy, Seaborn, Scikit-learn), SQL, R.
- **Machine Learning:** Supervised Learning (Regression and Logistic models), Unsupervised Learning (Clustering, Dimension Reduction, Abnormal Detection), Ensemble Learning (Random Forest, Gradient Boosting), Deep Neural Network (TensorFlow, Pytorch, FastAI).
- **Statistical Inference:** Hypothesis Testing, Causal Inference, Uplift Modelling (Bayesian MMM, Difference in Difference, Meta-Learners).
- **Generative AI & LLMs**: Stable Diffusion, Prompt Engineering, RAG, PEFT, HuggingFace, Agentic AI (LangChain, LangGraph, AutoGen, LlamaIndex, Langfuse).
- **Cloud & MLOps**: GCP (BigQuery, Vertex AI, Cloud Build), AWS (Bedrock, SageMaker, OpenSearch), Docker, GitHub.
- **Data Platforms**: Redis, Pinecone, ChromaDB.
- **Collaboration & Agile**: Bitbucket, Confluence, Kanban Boards.

PROFESSIONAL EXPERIENCE

WooliesX - Woolworths Group, Sydney, Australia

Data Scientist | August 2021 – Present (4 years)

- **Bayesian Marketing Mix Model & Paid Search Optimisation**
  - Developed a Bayesian Hierarchical MMM with MCMC. Used Adstock and Hill transformer to integrate lagged media impact. Gave measurement reports with quarterly frequency for $14 million Total Revenue for different Paid Search Channels, with example ROI around 13 for Performance Max.
  - Built an optimisation pipeline based on Gradient Boosting to reallocate Paid Search Shopping Ads budgets across 6 categories with 20k products, improving related ROAS from 9.7 to 10.8 and reversing the weekly decreasing trends of Paid Search click number from -8% to 12%. Built a CI/CD pipeline with GitHub and GCP Cloud Builder to give weekly results.
- **Everyday Rewards Member and All-Customer Campaign Performance & Uplift Modelling**
  - Created a Difference-in-Difference pipeline to measure product-level sales uplift for the Everyday Rewards member campaign. Corrected the overestimation from the previous pipeline by 23% and weekly measured around $400k incremental sales. The pipeline was soon welcomed by broader stakeholders and extended to All-Customer category-level campaigns in 2024.
- **Everyday Rewards Member Engagement Scoring Engine**
  - Designed a hierarchical Meta-Learner Regression Pipeline to quantify Everyday Rewards member engagement across the Woolworths ecosystem, including partners such as PetCulture, HealthyLife, etc. Generated monthly reports for 14 million Everyday Rewards members. This pipeline has been used in Customer Segmentation and Personalisation Targeting Models to guide the Everyday Rewards growth strategy until now.

Data Scientist with Contract | October 2020 – August 2021 (9 months)

- **Everyday Rewards Member Segmentation via Shopping Behavioural Clustering**
  - Built a robust clustering pipeline using Yule's Rule and Gaussian Mixture Models to segment Everyday Rewards members by cross-product spending behaviour. Gave not only segmentation but also soft features for 8 million active Everyday Rewards members. It has been widely supporting almost all Everyday Rewards campaign squads and membership insights projects since then.

The Australian National University, Canberra, Australia

Tutor | February 2020 – June 2020 (5 months)

- **Course Support & Assessment**
  - Delivered weekly workshops and tutorials.
  - Guide undergraduate computer science students for assignments and exam preparation.
  - Provided assessment feedback and exam grading for AI and machine learning statistical units.

Machine Learning Engineer (Research Assistant) | July 2019 – November 2019 (5 months)

- **Computer Vision Pipeline for Ant Head Orientation**
  - Designed a CNN-based Deep Learning Model with TensorFlow and Keras to predict ant head orientation in biological footage. Transformed it to a full pipeline with GUIs for non-technical users. Improved the prediction accuracy from 4% to 87% across diverse video backgrounds.

PROJECTS

- **Generative AI Document Assistant via LangChain Agentic RAG**
  - Built a LangChain and Langfuse PDF Reader web application combining RAG architecture with Pinecone, Redis, and ChromaDB. Enabled real-time modular pipeline updates and integrated human-in-loop with user feedback to optimise the module combination.
- **Generative AI Agent via OpenAI SDK from Scratch**
  - Built an agent based on OpenAI SDK to answer customers' questions related to the developer's career. Note down the message history and push notification to the developer's mobile phone.

EDUCATION

- **Master of Computing (AI Specialisation) | Australian National University, Canberra, Australia**
  Grade: Distinction (6.5/7)
- **Bachelor of Engineering (Information Engineering)| Beijing University of Posts and Telecommunications, Beijing, China**
  Grade: High Distinction (85/100)

INTERESTS

Old-style movies, pop music with unfamiliar languages, swimming, and hiking.

References Available upon Request