Kaihui Xie

(510) 710-3028 | kaihui_xie@berkeley.edu | San Francisco Bay Area, CA | LinkedIn: linkedin.com/in/kaihui-xie

EDUCATION

University of California, Berkeley, Berkeley, CA 08/2023-08/2024
Master of Analytics, **GPA**: 3.88/4.0
Relevant Courses: Database, Machine Learning, Deep Learning, Supply Chain, Financial Engineering, Optimization

Southeast University, Nanjing, China 2019-2023
Bachelor of Management, Electronic Business, **GPA**: 3.72/4.0
Relevant Courses: Data Structures, Operations Research, Data Mining, Big Data Analysis, Statistical Inference, UI, MIS

SKILLS

Programming Languages: Python, JavaScript, C, SQL, Java, R
Web Development: Flask, Node.js, HTML, CSS
Machine Learning Frameworks: PyTorch, TensorFlow, Scikit-Learn, Keras
Databases and Framworks: MySQL, Microsoft SQL Server, MongoDB, Google Cloud
Data Analysis Tools: Pandas, Matplotlib, Seaborn, Excel, Tableau
Version Control and Automation: GitHub, Docker
Languages: Chinese, English, French

INTERNSHIP EXPERIENCES

Machine Learning Engineer Intern - *Silicon Valley Commerce, Berkeley, CA* 05/2024-present
- Developed a web scraping tool using **Tampermonkey, Python, and MySQL** to automate data collection from e-commerce platforms, significantly reducing the manual research efforts
- Built an automated platform using **LLMs** (**GPT-4o**, **Stable Diffusion**) to generate A+ content for Amazon listings and leveraged **Flask** and **Node.js** to design the back-end infrastructure, increasing conversion rates by 10%-25%.
- Utilized **GitHub** for version control and collaboration, ensuring smooth project management and teamwork
- Performed **prompt engineering** and **fine-tuning** techniques to the automatic A+ content generator, resulting in a 45% increase in graphic designer and quality control approval rate
- Conducted **A/B testing** to ensure the robustness and accuracy of deployed models

Consulting Intern, HR Consulting (Automobile team) - *Mercer, Shanghai, China* 06/2022-08/2022
- Developed and implemented a data pipeline using **Python** to screen compensation data and perform quantile regression calculations, significantly improving the precision of compensation strategies
- Conducted a meticulous data cleansing process, identified and corrected over 200 errors in compensation surveys, enhancing data accuracy by 15%
- Participated in data preparation and matched over 20,000 position titles for automobile industry talent research projects
- Drafted a 22-page industry report on talent trends influenced by electric vehicle and autonomous driving technologies

Intern, Financial Services Department - *China Galaxy Securities Co.Ltd., Nanning, China* 07/2021
- Automated short-term stock selection by innovatively proposing and creating a **Python**-based application using Tencent Finance API, greatly improving stock selection efficiency and accuracy, available on GitHub
- Created a user-friendly interface with **TkInter** for the stock selection tool, facilitating non-technical user interaction
- Assisted in creating investment strategy visualizations by producing 5 detailed charts and reports, used to communicate key findings to stakeholders, reinforcing customer relations and data-driven decision-making

PROJECTS / RESEARCH

Predictive Analytics for Soccer Performance (Python) 01/2024-05/2024
Analytics Lab Course Project, University of California, Berkeley
- Developed a robust data pipeline and utilized **OOP** principles in **Python** to acquire and process soccer event data sourced through the Statsbomb API for machine learning model training
- Performed feature engineering and used **PyTorch** framework to implement ensemble learning with algorithms including Random Forest, XGBoost, and CNN to predict expected goals and passing success, improving recall by 15%
- Created an interactive website using **Streamlit** that enables users to upload event data and explore detailed performance analyses for teams and individual players, with customizable metrics for performance tuning

Aspect-Based Analysis of User Reviews' Impact on Online Healthcare Consultations (Python) 12/2022-05/2023
Graduation Dissertation, Southeast University
- Scraped 5M+ consultation records and 1M+ reviews from an online healthcare APP and removed irrelevant symbols
- Utilized LCF-ATEPC, an aspect-based sentiment analysis method integrated with a domain-adapted BERT model, to extract aspect words and predict their sentiments
- Clustered 14 categories of synonymous aspect words as impact factors using K-Means and analyzed the impact of each category on doctor consultation using linear regression