Kaihui (Theo) 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: Python (PyTorch, Scikit-Learn, Keras, Pandas, Flask), JavaScript, C, SQL, Java, R
Data Analysis Tools: MySQL, MongoDB, Spark, Tableau, Power BI, Excel
Languages: Chinese, English, French

PROFESSIONAL EXPERIENCE

Machine Learning Engineer (E-commerce Platform) - *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, reducing manual data entry by 80% and enabling reliable insights for operational decisions
- Built an automated A+ content generation platform using LLMs (GPT-4o, Stable Diffusion) to improve Amazon listing performance and leveraged Flask and Node.js to design the back-end infrastructure, resulting in a 10%-25% increase in conversion rates and used A/B testing to evaluate the impact of generated content on customer engagement
- Conducted exploratory data analysis to fine-tune content quality and increase graphic designer and quality control approval rate by 45% through prompt engineering and content quality adjustments
- Enhanced project collaboration through GitHub for version control, improving workflow efficiency by 20%

Marketing Analyst Intern - *Kantar, Beijing, China* 09/2022 – 11/2022
- Designed and implemented a nationwide survey to evaluate marketing campaign effectiveness for an online healthcare platform, focusing on key performance metrics such as brand awareness, view-through rate, and user intent
- Performed analysis on campaign performance, identifying trends in ad engagement, brand perception, and user intent
- Developed a 26-page data-driven report with clear visualizations and actionable recommendations, guiding the client to optimize campaign strategy and enhance future targeting efforts

HR Consulting Intern (Automobile team) - *Mercer, Shanghai, China* 06/2022-08/2022
- Developed a data pipeline with Python to screen compensation data and perform quantile regression analyses, improving compensation strategy precision
- Built a compensation analysis dashboard in Excel to visualize trends for Mercer's automotive clients, enhancing transparency in reporting and improving client satisfaction by 10%
- Conducted data cleansing and validation on 20,000+ data points, correcting 200+ errors, which increased the data accuracy rate by 15% for effective analysis in compensation projects
- Participated in an industry-wide analysis project, leveraging statistical tools to evaluate talent trends for electric vehicles and autonomous driving, and drafted a 22-page report for executive insights

Financial Analyst Intern - *China Galaxy Securities Co.Ltd., Nanning, China* 07/2021-07/2021
- Built a Python-based stock selection tool to automate daily stock recommendation report generation, increasing stock selection accuracy by 30%, reducing manual effort by 50%, and saving analysts time on deeper research
- Applied ARIMA models to forecast stock price trends, enhancing the accuracy of stock recommendations
- Created visualizations using Excel to convey key investment insights to clients, improving customer understanding

PROJECTS / RESEARCH

Predictive Analytics for Soccer Performance 01/2024-05/2024
Analytics Lab Course Project, University of California, Berkeley
- Designed and implemented a data pipeline to collect and process soccer match data, performing EDA to identify key metrics for predictive models on match outcomes
- Performed feature engineering and used PyTorch to implement ensemble learning with algorithms including Random Forest, XGBoost, and CNN to predict expected goals and passing success, improving recall by 15%
- Developed an interactive dashboard with Streamlit, allowing users to upload event data and gain insights into team and player performance, customizing metrics for detailed analyses

Aspect-Based Analysis of User Reviews' Impact on Online Healthcare Consultations 12/2022-05/2023
Graduation Dissertation, Southeast University
- Conducted aspect-based sentiment analysis on 5M+ consultation records, using EDA and clustering to categorize and understand customer satisfaction and its impact on doctor consultation
- Implemented a domain-specific BERT model for aspect extraction and sentiment prediction, yielding actionable insights for business strategy adjustments in doctor service quality and customer retention