Rucheng Zhou

+1 3238189639| ruchengz@usc.edu| ruchengzhou.com|github.com/zhouruc16

EDUCATION

University of Southern California Los Angeles, CA
Master of Analytics(MS) 01/2022–12/2023
University of Rochester Rochester, NY
Applied Mathematics (BS) Minor: Statistics, Studio Arts 06/2016–05/2020

SKILLS

- **Analytical Tools**: Python(Pandas/Sklearn/PyTorch/TensorFlow), **SQL** (MySQL/Hive/NoSQL), AWS, Tableau, MLops
- **Data Modeling**: k-means /XGBoost/LightGBM/BiLSTM-CRF/BERT/XLNet/NLP/GNN/LSTM/RNN
- **Certification**: AWS Machine Learning Specialist(2023)

WORK EXPERIENCE

Data Scientist, Big Data Team 09/2020–12/2021
China Construction Bank Shenzhen,China

Summary: Led a team of 3 to develop "Hui Understand You", a streamlined digital microloan platform. Spearheaded an end-to-end automated user financial behavior extraction, processing, integration, visualization and modeling systems in **pipelines**, enhancing the efficiency of risk assessment and optimizing user engagement and loan approval rate by 26.7% after launch.

- **Platform Development:** Pioneered the development of "Hui Understand You" and took charge of designing the comprehensive development strategy and execution guidelines, especially the **end-to-end automated data analysis system**; Delivered phase-specific analysis reports to stakeholders, ensuring precise oversight, pacing, and management
- **Behavior Analysis:** Leveraged **Tableau** to develop user behavior dashboard to recognize and track abnormal patterns, such as sudden spikes detection, unexpected payment exploration, and triggered alerts for further investigation automatically
- **Network Mining:** Utilized passive user data in conjunction with **graph-based models(GNN)** to conduct network analysis, enabling precise identification and clustering of fraudulent and default patterns; Provided valuable insights from comprehensive investigation to optimize and fine-tune anti-fraud strategies
- **Risk Assessment:** Refine the credit scorecard model in fusion with customized **LSTM** model and anomaly detection techniques, elevating model prediction precision by 17% offline and loan approval rate by 26.7% after launch

RESEARCH

Direcred Researcher - AI chatbot development 01/2024–Present
University of Southern California Los Angeles, CA

• Developed a chatbot called Coursistant, using **custom embedding model and GPT3.5** turbo to answer course material and logistic questions, enhancing student engagement and support.
• Utilized **llama-index** library for creating document embeddings of Piazza Q&A and video lecture transcripts, enabling efficient information retrieval.
• Integrated with a third-party Piazza API to automate responses to new questions posted on Piazza, with an hourly update mechanism for embedding new Q&A pairs.

PROJECT EXPERIENCE

Empowering User Experience - Optimizing Product Recommendations through AB Testing and UI Enhancements

• Conducted comprehensive data analysis and collaborated with UX designers to develop refined UI prototypes based on data-driven insights; Employed rigorous statistical methodologies to ensure the experiments were well-structured, randomized, and statistically significant, providing actionable results
• Developed and fine-tuned a sophisticated Logistic Regression model to assess the probability of successfully launching the UI changes; Delivered in-depth customer insights by interpreting the model results, enabling data-driven decision-making for the product development
• Led the end-to-end UI change implementation process, starting from identifying improvement opportunities through extensive data exploration and user feedback analysis; Continuously monitored key performance metrics, ensuring successful tracking of product usage rate and user behavior post implementation

News Intelligence - Unleashing the Potential of User History for Click Prediction

• Led a dynamic team in the development of a cutting-edge news r**ecommendation system**, harnessing user historical browsing and click data to predict future click behavior accurately
• Leveraged the power of Python to conduct comprehensive data mining on more than 200,000 news app users, encompassing click environment, click volume, news co-occurrence frequency, article length, and click preferences.
• Completed a multi-way combination of recall strategies, expertly incorporating item-based collaborative filtering (itemcf), embedding techniques, and **YoutubeDNN** to **enhance click predictions**
• Selected and fine-tuned three representative ranking models, including **LGB Ranker, LGB Classifier, and DIN** (Deep Interest Network), achieving top-notch performance in click prediction;