Shijie (Kirk) Han

(332) 260 8224 ¦ sh4460@columbia.edu ¦ linkedin.com/in/shijie-kirk-han

EDUCATION BACKGROUND

Columbia University   **New York, US**
Master of Science - Operations Research (3.6/4.0)   Sep 2023 – Dec 2024
University of Liverpool   **Liverpool, UK**
Bachelor of Science - Mathematics (4.0/4.0)   Sep 2019 – Jul 2023

PROFESSIONAL EXPERIENCES

Alibaba Group   **Beijing, CN**
Data Scientist Intern   Jun 2023 – Sep 2023
- Crafted scripts for tracking CRM system user interactions and fulfilled data extraction and log reports.
- Adopted funnel model to monitor user behavior trajectories, emphasizing significant data transitions and pivotal engagement junctures, enabling more targeted approach to user interaction and retention.
- Applied cohort analysis triangle to unravel intricacies of both product and user life cycles, refining and bolstering customer retention strategies through deeper understanding of user engagement patterns.
- Integrated RFM model with K-Means clustering to achieve granular user segmentation, which informed development of customized promotional strategies, enhancing user experience and conversion rates.

Nestle   **Beijing, CN**
Data Scientist Intern   Jun 2022 – Sep 2022
- Optimized storage and cleansing scripts, increasing data transformation efficiency to ODS layer by 50%.
- Conducted AB-Testing on new application click buttons to assess their impact on the browsing patterns of target user groups, offering robust data-driven insights for new product iterations.
- Implemented XGBoost algorithm to analyze the application browsing patterns of more than 100k brand members, using SHAP to visualize feature contributions which accurately evaluated the impact of users' key behaviors on product choices, aiding 30% increase in personalized marketing success.

McKinsey & Company   **Beijing, CN**
Business Analyst Intern   Jun 2020 – Sep 2020
- Designed interview guidelines and coordinated cross-departmental workshops; Responsible for compiling discussion content into reference documents, supporting team decision-making and project progression.
- Assisted clients in optimizing media marketing strategies and user persona analysis systems, enhancing content recommendation quality, facilitating personalized marketing and digital transformation.
- Synthesized diverse information sources for comprehensive competitive analysis of clients, adeptly pinpointing product strengths and weaknesses through gathering pertinent industry data and examining rival marketing tactics and crafted engaging narratives to reveal practical strategic insights.

PROJECT EXPERIENCES

Credit Card Fraud Detection based on Logistic Regression   Sep 2023 – Dec 2023
- Processed raw data through label smoothing, categorical variable numeration and numerical variable standardization, boosting model's predictive accuracy; Utilized SMOTE oversampling technique to balance dataset, synthesizing minority class (fraud) samples to enhance model's ability to detect anomalous activities.
- Employed logistic regression model with cross-entropy loss, training with varying regularization strengths; Selected optimal parameters based on average recall rates from cross-validation; Applied model in test set and adjusted threshold to plot confusion matrix, achieving 0.93 AUC score in final fraud detection test.

Sentiment Analysis based on LSTM Model   Oct 2022 – Dec 2022
- Utilized pre-trained GloVe vectors as reference, extracting words to form index matrix to connect lexicon.
- Created normalized batch dataset with labels and set hyperparameters for constructing stacked LSTM network. Applied Dropout and early stopping to mitigate overfitting. Selected cross-entropy loss and Adam optimizer for optimal performance, culminating in a 92% analytical accuracy on the test dataset.

TECHNICAL SKILLS

Python | Pandas | Numpy | Matplotlib | PyTorch | TensorFlow | Sklearn | Gurobi | SQL | Spark | Tableau | Power BI | AB-Testing | Random Forests | GBDT | XGBoost | LightGBM | SVM | CNN | RNN | LSTM | Transformer