



Kai Yang

EDUCATION

East China University of Science and Technology (ECUST) Master of Engineering Management 2019.07-2022.05

- **Main courses:** Project Engineering, Management, Operations Research, Business data analysis, Business intelligence;
- **Graduation project:** Based on demand management and progress management to solve the construction and research of T company's smart travel route recommendation service, using market research and software development management, product finally achieve the top3 profit share in the OTA market;

Xi'an Jiaotong Liverpool University Financial Mathematics BA 2011.09-2015.07

- **Main courses:** Interest Theory, Corporate Finance, Financial Accounting, Securities Market, Risk Theory, Game Theory, Corporate Finance, Macro and Microeconomics, Linear Algebra, Calculus;
- **Graduation project:** Use R to build a time series GARCH model, fit the CSI 300 stock index futures trend model, and verify the importance factors in the arbitrage strategy;

WORK EXPERIENCE

Tongcheng Travel Holdings Limited Recommendation Algorithm Engineer 2020.7-2022.7

- **Artificial Intelligence Customer Service:** Through user intention identification and understanding, word segmentation and sentiment analysis are performed based on basic features in historical data. For single-round QA and simple Duolun Task tasks, build an OTA industry knowledge base and label system, use FastText and Bert offline training to achieve an accuracy rate of more than 90%, perform Query retrieval and recall for different user intentions, and pass the possibility of candidate intentions. Higher reordering fusion rules increase the matching of cold starts and high frequency problems.
- **HuiXing cross-recommendation system:** Optimize the revenue and user conversion rate in the transit route, use the combination of XGB and deep learning to solve the pain points of users' travel difficulties during peak periods and low success rate of ticket grabbing, and recommend various types train plus train etc. for users under different search categories. Combination of multiple routes such as train + plane, increase user click rate and order rate;
- **UGC Community:** Build the UGC content platform of Tongcheng Mini Program from scratch, start from tag annotation and user travel records, identify travel boutique stickers in the early stage and distribute them to the corresponding city and age group user groups, dynamically calculate user interests through MAB and recall the most suitable content, use the generalization index of each stage to train the DNN model under multi-objective to solve the content distribution and sorting;

OYO Hotel(Masayoshi Son invest) Senior Data Mining Engineer 2019.04-2020.03

Construction of risk control platform

- Create 500+ features to judge fake transactions by hotel and order dimensions, including hotel owners' equipment features, order GPS and IP features, behavior features of users placing orders on the day, owner's PMS operation trajectory on the day, and time-sharing operation frequency in different time windows etc., to create a fake hotel transaction model (including brushing and missing orders) to help the company save thousands of orders every day to make up for fraud;
- According to the ResNet model, it only costs one camera in each hotel to monitor and identify the number of people who deliver ID cards back and forth at the hotel front desk every day in real time, to compare the number of orders entered by hotel owners on the PMS platform, and to further accurately identify false hotel transactions level, and make reports to visualize and analyze abnormal fluctuations of hotel OCC indicators;

Coupon function analysis(referral traffic)

- Add user imei identification, verification code page operation and click behavior, model matching, IP fluctuation positioning, GPS black product aggregation, startup and running time, add unsupervised model on the basis of lightgbm, filter features based on varclustering, and generate iforest tree model Identify the results, intercept 20-40% of registered black products every day;
- Create a fake coupon transaction model and a dynamic coupon issuance model. There are two types of models: new customers and regular customers. The regular customer model is mainly used to predict whether to place an order again. Confirm the amount of coupons issued for the next event and whether there is a threshold. The bottom layer of the two types of models uses LSTM and XGB to output normalized scores, and the upper layer uses LR to improve interpretability. After going online, increase the conversion rate of orders by 10%, and the order rate of customers who issue coupons is high;

Shanghai Kuai Niu Group(Jing Dong invest) Head of Data Analysis 2018.04-2019.03

Product function innovation

- **Recommendation ranking:** Using the user's app installation list and the demand preferences of different loan products, according to the historical user's click habits, using clustering and classification algorithms, draw different portraits of users to improve the user's product purchase conversion rate by 20%;
- **Social network:** For the first time in our company, we use the app tracking point and user operator data to code 0 and 1 for users through the data funnel method, and then add the A-card score. Each time, the user's social score is updated according to the weight of the social network. User social risk system, and add B card and C card as features, about iv0.1;

Pre-loan score card

- Analyze and mine user behavior data, loan short message data, collection record data, operator detailed bill data, mobile phone address book data, etc., and derive 100+ features;
- Responsible for the design, construction, optimization and integration of Kuai Niu risk control data warehouse, including demographic characteristics of new and old customers, tripartite attributes, behavior data and loan data, a total of four modules, 10 tables with over 2000 fields wide, Greatly improve modeling efficiency;

Everbright Bank anti-fraud program

- Responsible for the anti-fraud project of China Everbright Bank, building a total of 15 wide sheets. Through preliminary analysis and cross-validation, 10+ feature multi-dimensional subjects are designed, 5000+ derived variables are created through FRM clustering, ylabel is established semi-supervised, and text-type features are classified;
- The XGB model reduces the total number of outgoing calls by 30%, improves the efficiency of Everbright's outgoing calls to determine fraud, the recall rate is 99%, and the false positive ratio of outgoing calls is controlled below 1:20;

Shanghai Nami Finance**Risk Control Data Analyst****2017.04-2018.04**

- Responsible for SQL fetching, variable ETLization, variable screening, and Python development of logistic regression models. A total of 5 models have been established, including pre-loan and post-loan scorecards. The pre-loan scorecard has undergone 3 iterations, and the self-packaged model cooperates with the big data interface. Online, the repayment rate is over 91%;
- Mining nearly 20 valid strong tags from third-party data sources, user text messages, and bank flow, and putting them into the pre-loan scorecard model, the overall AUC of the pre-loan scorecard is increased by 1%, and the ks value is increased by 2%;

PERSONAL SKILL

IT skills: Master Linux, Python, use Mysql, MongoDB and Spark tools proficiently, and be good at Hive database query, extraction and processing.

Hobbies: Basketball, badminton, reading