Yujie (Leo) Li

yujie.li.2020@anderson.ucla.edu | 217-377-7722 | linkedin.com/in/yjl96 | Authorized to work in the U.S. permanently

EDUCATION

UCLA ANDERSON SCHOOL OF MANAGEMENT   *Los Angeles, CA*
Master of Science in Business Analytics (GPA: 3.86/4.0)   *Expected Dec 2020*
- SQL and Data Management, Machine Learning, Prescriptive Modeling (A/B Testing), Data Visualization, Data Storytelling, Optimization, Customer Analytics, Competitive Analytics, Operations Analytics

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN   *Champaign, IL*
Bachelor of Science in Finance (GPA: 3.70/4.0, Honored, CFA level 1)   *May 2019*

TECHNICAL SKILLS

- Programming: SQL, R, Python (Pandas, Numpy, Scikit-learn, Tensorflow, Matplotlib, Seaborn)
- Software: Tableau, Adobe Analytics, Snowflake, Fivetran, DbSchema, Gurobi, MySQL Workbench, Excel (Solver)

DATA SCIENCE PROJECTS & CHALLENGES

SQQUID E-COMMERCE DATA ANALYTICS   *Mar 2020*
- Designed a data pipeline to process Sqquid OLTP database through Amazon Web Services and loaded more than 500K+ selling and shipping records from online public database into private data warehouse Snowflake data lake with streaming tool Fivetran
- Connected schemas from database to Tableau to retrieve insights on fulfilments and sales and found out Top 10 stores with longest/shortest fulfilment days, customer cancellation rate, and orders out of stock by MySQL Workbench
- Analyzed the delayed delivery problem for certain stores and offered four recommendations to shorten lead time

CUSTOMER DEMOGRAPHIC MARKETING STRATEGY A/B TESTING PROJECT   *Jan 2020*
- Performed A/B Testing to measure the impact of coupons that were sent to customers randomly and determined which customers to target based on demographic and behavioral information
- Regressed treatment (coupons assignments) on age dummies and past revenue to confirm that treatment was assigned randomly
- Analyzed omitted variable bias from simple regression that regressed revenue on treatment, age dummies, and past revenue; improved the model by including interaction terms and concluded from the coefficients and significance that consumers with intermediate levels of past revenue are the most attractive in terms of targeting

BANK MARKETING CAMPAIGN MACHINE LEARNING PROJECT   *Nov 2019*
- Built a responsive model with two Machine Learning classifiers (Logistic Regression, Random Forest) to label potential customers and tuned parameters to evaluate the effectiveness of a marketing campaign launched by a bank
- Compared the accuracy of results among different methods by the confusion matrix and ROC curve; selected the model by accuracy, precision, and recall rate to target potential customers in the future market; tuned the model by Grid Search to choose the optimal hyperparameter values
- Interpreted the results to show how targeting the selected customer segment could benefit the bank and increase bank's revenue

ADOBE ANALYTICS DATA ANALYSIS AND VISUALIZATION CHALLENGE 2019   *Sept 2019*
- Analyzed and decomposed problem as a leader to perform an explanatory marketing analysis for MLB's website and mobile apps
- Conducted data mining on ads clickthrough rate, conversion rate, churn rate, marketing channel, ticketing, online purchase, and other KPIs to identify customer behaviors of different user cohorts; plotted linear regression lines, histograms, bar/area/flow/pie charts to visualize the results with Adobe Analytics
- Defined churn rate and conducted hypothesis on factors that affect churn rate; provided recommendations to test the hypothesis by launching a difference in difference model on an experiment to redistribute marketing resources to Facebook mobile app by 10 more percent instead of paid search for better advertising effectiveness

PROFESSIONAL EXPERIENCE

SPORTSDT (Leading sports data provider in China)   *Shantou, China*
Machine Learning Analyst Intern   *May 2019 - Aug 2019*
- Analyzed over 4000 English Premier League pre-game insights from experts on betting strategies and built a random forest model based on the word matrix and gambling odds to predict game results
- Cleaned text data and selected 50 key words from comments using term frequency - inverse document frequency rating (TFIDF), a Natural Language Processing method to rate the words in a passage, to build a 50-dimension word matrix
- Tuned the parameters of the model and gained 54.37% accuracy rate on the test set, which beat the purely objective probability

ABBOTT   *Shanghai, China*
Planning Team Intern   *Jun 2016 - Aug 2016*
- Conducted feasibility study of changing shipping method of products from ocean to railway, potentially reducing transit time from Europe to China by 18 days on average
- Evaluated the tradeoff between overall profitability due to shorter inventory turnover and reduced holding cost vs. the increased cost of transportation; determined that the profitability of reducing lead time is proportional to the size of products
- Recommendation was adopted by vice president to potentially improve the profits of 32 infant formula products