Han Gong

217-979-9875 | https://www.linkedin.com/in/han-gong-20237214a/ | han.gong@duke.edu

EDUCATION

Duke University May 2022 (expected)
Master of Science, Interdisciplinary Data Science
Coursework: Machine/Deep Learning, Statistical Modeling, Bayesian, NLP, Data Engineering System, Algorithm

University of Illinois at Urbana-Champaign (UIUC) May 2020
Bachelor of Science in Statistics & Advertising

TECHNICAL SKILLS

Programming: Python(scikit-learn, Keras, TensorFlow), R, SQL, SAS, Hive, HTML
Analytics & Skills: AWS(SAA), SPSS, Tableau, Google AdWords/Analytics, Flask, BigQuery

EXPERIENCE

Lyft Inc.
Data Science Intern San Francisco, CA | May 2021 – Aug 2021
- Explored behavioral traits of tiered drivers in the driver loyalty program, identified patterns of tier distributions in different markets, and built a classification model to predict driver tiers with early indicators
- Collaborated with cross-functional partners, estimated the incremental gain of the referral campaign for high-value drivers with hierarchical models, and delivered experiments decision with ROI analysis
- Updated stratification script using clustering method to improve the efficiency of driver experiment

NetEase Cloud Music
Data Analyst Intern Remote | April 2020 – Aug 2020
- Conducted three AB testing for a new user module on the homepage, collected and analyzed data for user growth, proposed suggestions on pop-ups to drive the click-through-rate by 5%
- Monitored user acquisition & retention data, developed HiveQL code and built a user-friendly dashboard to visualize daily registration/login funnel from mobile-end, third-party binding and recall users data
- Evaluated user behaviors through attribution analysis, providing data support to improve hot song recommendation algorithm and user reward mechanism

Technology and Social Behavior Lab
Research Assistant Urbana, IL | Sept 2018 – Jun 2020
- Conducted 2x2 between-subject study "Persuasion Effects of Message Evidence under Different Modalities", applied regression model and paired t-test on SPSS to detect the interactive effect
- Collected and validated data for ongoing research on "Decision Making Under Surveillance", extended the study results in the application of target advertising under datafication

PROJECT HIGHLIGHTS

Fake News Detection with LSTM and Word Embeddings *Fall 2020*
- Extracted and tokenized word features from 10,000+ labeled news statement using NLTK library, constructed word embeddings of GloVe and Word2Vec to perform similarity task for insights
- Implemented LSTM model for fake news classification using Python with 70%+ accuracy

Obesity Level Estimation Based On Eating Habits and Physical Conditions *Fall 2020*
- Defined and finalized variables based on previous researches on obesity, accessing the impact of 17 lifestyle features and their interaction effect using proportional odds, multi-nominal and SVM model
- Applied model fit testing and statistical inferences to evaluate model performance

Walmart Predictive Modeling Project *Fall 2019*
- Generated exploratory data analysis, applied feature engineering and developed machine learning code on over 3 million observations to predict sales of 111 weather-related items in 45 Walmart stores
- Deducted 30% RMSLE by correcting skewed data through transformation and classification