

XI 'CASSIE' GUO

Denver, Colorado 80237
(952) 465-6453 guoxx456@umn.edu github.com/lilsummer
T-shaped Data Scientist with a passion for data-driven business solutions

EXPERIENCE

Data Scientist – Travelport (Denver, CO) January 2018 - Present
Machine Learning Research and Prototyping

- Conversion rate prediction over flight search response using GBDT + LR stacking model; GPU utilization and recall optimization (10 min online prediction over 2TB of hourly data)
- Price elasticity estimation for key markets (statsmodel, dask, pyspark)
- Deep reinforcement learning framework for pricing machine (revenue management + openAI gym)
- Metric design to find cheaper split tickets to save 20% internal production cost (ranking algorithm)

Production Support

- Partnered with data engineers to implement Spark streaming in production enabled 25G of daily data transferring and feature engineering (pyspark, Hive, SQL)

Business Function

- Led 6 monthly dynamic pricing research group meetings and published 6 internal articles

Data Science Intern - Metabiota (San Francisco, CA) June 2017 - August 2017
Parsimonious representation of global influenza pandemic scenarios

- Constructed data pipeline to process 12 TB of raw data using R on AWS (Sparklyr, data.table)
- Defined key metrics and built GLM models to access the pandemic risks of client interests (H2O)

Data Science Fellow - The Data Incubator (San Francisco, CA) March 2017 - May 2017

TECHNICAL SKILLS

Programming: Python, R, scala, SQL, Shell.

Tools and Platforms: Spark, Hive, H2O, Azure, AWS, GIT, Docker, Tableau, Plotly.

SELECTED PROJECTS

Risk Factor Analysis on the Impact of Climate on Disease Transmission (non-parametric statistics)

- Worked with a software engineering team and launched a mobile survey app to collect key risk factors in real time; used by more than 40 farms to prevent 30 million economic loss
- Pinpointed 56 years of weather data from NOAA; tracked the risk of weather indicators by non-parametric tests; results were utilized for state-wide biosecurity guideline recommendation
- Extracted location features from eBird reports; predicted the trend of index cases at 95% coverage using ARIMA, GLM, and GAM models

Smart Fitness Advisor (R, Shiny)

- Prototyped three ML models to address different health concerns based on 5 years of CDC survey data; Logistic regression model classified outcome at 91% precision rate
- Created a Shiny App to give crowd-sourced health recommendations based on personal biometrics

EDUCATION

University of Illinois, Urbana-Champaign June 2019 - November 2020
Master of Computer Science (part-time/online)

University of Minnesota, Twin Cities, MN August 2012 - October 2017
Ph. D. of Epidemiology, graduate minor in computer science

China Agricultural University, Beijing, China August 2007 - May 2012
Bachelor of Science in pre-Veterinary Medicine

COMMUNITY AND LEADERSHIP

Denver Chapter Organizer, Data for Democracy, 2018