YAJING GAO
DATA SCIENTIST

Summary

Data Scientist with a Ph.D. in Biomedical Engineering who is interested in integrating knowledge from science, math, and engineering to solve problem in novel and efficient ways. Experienced in general programming and algorithms, statistics and analysis, data science models and visualization.

Experience

Metis — San Francisco, CA
Data Science Fellow — Jan 2017 to Current
- Developed several Data Science projects over the course of a 12-week immersive bootcamp, incorporating skills such as data wrangling, web scraping, regression, database, machine learning, and natural language processing.

Duke University — Durham, NC
Graduate Research Assistant — 2009 to 2016
- Thesis Topic: Mechanistic Models of Anti-HIV Drug Delivery.
- Used models based on physical processes to simulate drug delivery and drug effectiveness of new, novel HIV prevention drug delivery products.
- Results used and interpreted in drug design, suggested dosage regimen, estimation of effectiveness, and analysis of variability in clinical data.

Durham, NC
Teaching Assistant - Statistics & Mathematics — 2007 to 2011
- Assisted with Statistics and Mathematics courses.
- Organized and assisted teaching labs and study sessions in Biomedical Engineering.

Projects

Visualizing and differentiating subway stations from MTA Turnstile data
- Filtered raw data from MTA website to give entrance and exit data for New York Subway stations.
- Visualized station throughput by animated graph of each station over time by changing dot size and color to match the log of total riders.
- Data for each station over time were analyzed via a Fast Fourier Transform to classify station type based on phase and amplitude of the frequency data.

Analyzing trends in domestic box office over the past 30 years
- Combined box office score for top 100 movies normalized values by Consumer Price Index and total US population show a declining trend in total sales over the past 10 years.
- A predictor of movie gross for the top 20 to 100 movies are used to estimate the top blockbusters for each year.
- Factors such as the economy and top blockbusters do not significantly contribute to this trend.

Classifying at fault drivers in fatal accidents
- Obtained data for all fatal crashes in 2015 from the National Highway Traffic Safety Administration.
- Used Random Forest Classifier to differentiate between at fault and not at fault drivers by age, gender, car make, model, and year.
- Results show no difference between makes of cars, and male drivers 20 to 30 are more likely to be at fault in a fatal accident.

Contact

- ✉ yajing@gmail.com
- 🌐 yajingg.github.io
- ☎ (919) 339-1119
- 📍 San Francisco, CA
- in yajingg
- ⌂ yajingg

Education

Duke University
Ph.D. 2016
Biomedical Engineering

Duke Kunshan Univerisity
2016
Duke-Tsinghua Machine Learning Summer School

Duke University
B.S.E. 2009
Biomedical Engineering, Mechanical Engineering, Chemistry, Mathematics

Skills

PROGRAMMING
Java
C++
Python
SQL
MATLAB
HTML

DATA SCIENCE
Data Analysis
Data Visualization
Machine Learning
Model Building

SCIENCE
Statistics
Differential Equations
Numerical Analysis
Drug Delivery
Fluid Mechanics
Thermodynamics