YIMENG ZHAO

Objective: Data Engineer/ Data Scientist/ Data Analyst

Email: yimeng_zhao@utexas.edu   Phone: (512) 713 – 9628   Github: https://github.com/adv34715

PROFILE

- **Data mining**/machine learning, as well as data visualization using Python and other tools.
- Experience in **database** management; good understanding of relational and No-SQL databases.
- Basic knowledge of **statistics**, including regression and Bayesian methods, and ability of using R and SAS to do regression analysis.

EXPERIENCE

Data Analysis Intern, IC2 Institute, Austin, TX;   June 2016 - December 2016
- Collected and visualized universities data to verify universities academic development stimulate Austin's economic development.
- Designed and built a longitudinal database for Austin's entrepreneurs and startups using MySQL Workbench.

Research Assistant, Lab of Image Info Processing, HFUT, China;   January 2015 - June 2015
- Researched on Defocus-Based Depth Cues Acquisition with Light-Field Camera using Matlab.

EDUCATION

- **M.S, Information Science, University of Texas at Austin; Austin, TX**   May 2017
  Specialization: Data Science, Data Mining, Database Management, Web Development.   GPA: 3.9
- **B.E, Electronic Technology, Hefei University of Technology; China**   June 2015
  Course: Algorithms and Data Structure, Pattern Recognition.   GPA: 3.2

PROJECTS

Student Paring Systems, PenPal Schools, Austin, TX;   January 2017 - Present
- Draw student's data from **PostgreSQL** using Python package psycopg2 and store into csv files.
- Train classifiers to classify students into active and inactive ones with cross-validation method.
- Give students recommendations for choosing group members and new online courses

Bank Telemarketing Success Prediction   January 2016 - May2016
- Selected 20 features from more than 50 features and applied **SVM**, **LR**, **DT**, **RF** classification models to predict a bank's telemarketing success on about 50, 000 records.
- Compared and evaluated the four classification models by the Area under an ROC curve and explored how 20 feature (numerical and categorical) of telemarketing affect the success of telemarketing.

Crowdsourcing Workflow for Generating Human Data   September 2016 - December 2016
- Built a pipeline for generating and exchanging data from medical image records between **MongoDB** and a crowdsourcing website to transcribe images.
- Analyzed data and designed an algorithm to improve the reliability of the data from crowdsourcing.

Data Analysis of DPLA   June 2016 - August 2016
- Retrieved data aggregated by DPLA via a RESTful API, aggregated and analyzed contributors data.

SKILLS&TOOLS

Languages: Python, MySQL, C++, JavaScript, HTML, CSS, Bootstrap, PHP, Unix commands.

Databases: MongoDB (pymongo), PostgreSQL (psycopg2), MySQL Workbench.

Softwares/ Tools: Jupyter Notebook, Github, Amazon Mechanical Turk, AWS, Photoshop, Qlik Sense, Microsoft Office Suite.

Design: Dashboards, Posters, Infographics, Slides.