Add: Brighton, MA Tel: +2024760123 | Email: sanwang@gwu.edu

San Wang

GitHub blog: https://san-wang.github.io/ • LinkedIn: linkedin.com/in/san-wang

Product-thinking data scientist with solid background in advanced mathematics, statistical and machine learning modeling. Proficient with Python, Tableau, MySQL, Tensorflow, Scikit learn, Git and Cloud. Passionate about taking a business problem and turning it into actionable output. Looking forward to transferring data insight into value at your company.

EDUCATION

The George Washington University DC, 01/2016-12/2017

M.S. in Data Science, GPA: 3.9/4.0

Sichuan University Chengdu, China, 09/2011-06/2015

B.S. in Mathematics, concentration in Statistics

WORK EXPERIENCE

Research Assistant (Blog) DC, 05/2017-12/2017

The George Washington University (GWU)

- Led 4 graduate students to analyze data scientists' demand across U.S. job market using web scraping, NLP
- Analyzed U.S. top 100 universities' curriculum design with department contribution using Tableau
- Developed overview and industry statistics pages in GWU Data Science program blog

Operation Assistant Remote, 04/2016-Present

Jingenius LLC

- Promote marketing for hellogwu.com website, inDC mobile app and WeChat Official Account through social media
- Help international students merge in new culture and coordinate culture events with local business

PROJECTS

Movie Recommendation System [Python] (Demo) MA, 03/2018-05/2018

- Developed a Flask web API with mySQL as database to provide live movie recommendations from scratch
- Provided personalized recommendation according to users' rating history using collaborative filtering methods
- Provided content-based recommendation based on 45,000 movies' overview, cast, directors and 270,000 users' ratings

Customer Value Analysis [Python] MA, 01/2018-03/2018

- Predicted valuable credit card customers based on their payment history, bill statement history and some background information by using 25,000 samples with more than 20 features
- Built pipelines to improved performance by automating PCA, tuning regularization parameters and model selection

Achievement: Gaining 81.88% accuracy when applying to 5,000 holdout samples

Master Thesis, German Traffic Sign Classification [Python] (Blog) DC, 02/2017-12/2017

- Classified 39,209 images in 43 categories using convolutional neural network
- Built data-reading pipelines to automatically preprocess 39,209 images in 43 folders for both Caffe and Tensorflow
- Visualized the change of parameters, kernels' pattern and feature maps through whole training process in Tensorboard
- Reported model performance in a non-technical friendly dynamic Tableau dashboard

Accomplishment: Gained 90.58% accuracy when testing on 12,630 images in Caffe and 95.42% in Tensorflow on 7841 images.

CERTIFICATION

MySQL for Data Analytics and Business Intelligence, Udemy 03/2018

CP100A: Google Cloud Platform Fundamentals, ROI Training 11/2016

ACTIVITIES

Regular Attendee, Data Science Meetup/ Machine Learning Society DC/ MA, 2016-2018

Volunteer, Habitat for Humanity LA, 03/2017

Leader, Data Science Program Student Council DC, 05/2016-12/2017

SKILLS

Software: Tableau, Python, mySQL, Tensorflow, Scikit-learn, Caffe, R, Google Cloud Platform, Git, Docker