SHU ZHANG

Data Analyst

(515) 715 3215



https://mlikelihood.github.io



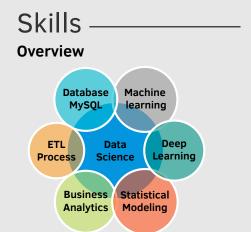
shuzhang@iastate.edu



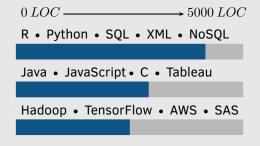
/in/shu-zhang-3ab863155



MLikelihood



Programming



Summary —

Interests - Data science and technology; Extraction of useful information **Accumulation** - Having implemented most of statistical models and machine learning models in R and Python; Proficient in data quality and security check and feature engineering; Good at model assumptions check, model evaluation and selection; Forecasts and visualization. Uniqueness - Studying Statistics and CS makes me juggle between the interpretability and accuracy of models, till find an optimal balance; Good model construction and programming skills.

Education

2016 - 2019 Master's Degree, Computer Science (Expected)

Iowa State University

2016 - 2019 Master's Degree, Statistics (Expected)

Iowa State University

Experience

2017 - Now **Graduate Assistant (Enrollment Research) Iowa State University** Project: Forecasting the enrollment of certain student groups at ISU

- · Constructed data set for time-series using feature engineering
- · Implemented multiple statistical and machine learning models
- Achieved high prediction accuracy and replaced the existing model
- Tools:R, Python,keras,scikit-learn,pandas,tidyverse,dplyr,ggplot2

Project: Strategic analysis and optimization of Awards-Structure

· Developed a simulation schema to maximize school enrollment & profit while keeping some other factors balanced (ongoing)

2016 - 2017 **Graduate Teaching Assistant (Statistics)**

Duties: Taught Business Statistics and Introduction to Statistics

Courses

CS: C Programming, Object Oriented Programming (Java), Machine Learning, Design & Analysis of Algorithms, Large Data Algorithm, Advanced Database System, Simulation: Algorithms and Implementations, Theory of Computation. STAT: R Computing, Excel Data Analysis, Probability Theory (I&II), Statistical Methods (I&II&III), Bayesian Statistics, Time Series, Survey Sampling, Advanced Statistical Methods

Projects

Dec 2017 -Jan 2018

Web Crawler and Page Rank

Large Data Algorithms

· Designed a web crawler that crawls pages from Wikipedia, identified the top k pages that are most relevant to the chosen topic by performing a weighted BFS and computed the page rank

Sep 2017 -Oct 2017

Topic sensitive hashing & Similar documents

Large Data Algorithms

 Estimated Jaccard similarity among documents by applying LSH and identified documents that have similarity more than 90%

July 2017 -Aug 2017

Zillow Home Price Prediction

- Created a weighted forecasting model from three models(XGBoost, LightGBM, and baseline) to predict house sale prices for Zillow
- Worked on multiple feature engineering using python(top 21%)

May 2017 -June 2017

Mercedes Benz Car testing time prediction

Kaggle Competition

• Predicted the time a car would take to pass testing process using an averaged model of Xgboost and a Stacking pipeline which consists of a cross-validated Lasso and a Gradient Boosting (top 18%)

Nov 2016 -Mar 2017

Models and optimization techniques

Machine Learning

- Implemented Naïve Bayes, Support Vector Machine, Neural Network, Random Forest, KNN, logistic regression and Ensemble models real datasets; Gained high validation and testing accuracy
- Learned optimization techniques such as Batch Gradient Descent, Stochastic gradient descent, regularization and weight decay