Guanyan Lin

1750 N Range Rd, IN, Bloomington
aaronarchiblad@gmail.com | 812-369-6456

EDUCATION

INDIANA UNIVERSITY BLOOMINGTON | MS in data Science
May 2018 | Bloomington, IN • Cum. GPA: 3.57/4.00

INDIANA UNIVERSITY BLOOMINGTON | BS in statistics | BA in Economics
Minor: Mathematcs, Business Consulting
May 2016 | Bloomington, IN • Cum. GPA: 3.34/4.00 | Major GPA: 3.74 / 4.00

WORK & RESEARCH EXPERIENCE

DANIEL MCDONALD'S LAB | Machine Learning Research Assistant
May 2017 – Present | Indiana Univeristy Bloomington, IN

- Read quantum chemistry papers and implement the most recent machine learning models and algorithms.
- Utilize programming languages, such as R and Python, to change the file format and clean chemical data.
- Use different molecule representations to transform input data such as BOB and Coulomb matrix.
- Implement kernel ridge regression and neural networks in R and Scikitlearn in serial and parallel.

INDIANA STATISTICAL CONSULTING CENTER | Statistical Anaylst Assistant
Sep 2015 - May 2016 | Indiana Univeristy Bloomington, IN

- Used strong analytical skill to assist graduate students and faculties by understanding their research projects and providing statistical solutions for research projects.
- Built statistical models such as generalized linear models and repeated measurement ANOVA to provide data driven evidences for research projects.
- Utilized programming software such as R, Excel, and Python to clean data and to run analysis on surveys and lab projects.
- Worked independently and collaboratively with 2 supervisors and 1-3 consulting associates to develop solutions for clients.

DATA SCIENCE PROJECTS

KAGGLE: House Prices: Advanced Regression Techniques
Oct 2016 - Nov 2016 | Bloomington, IN
Techincal skills: R • Excel • Log Transformation • Reguralized Linear Regression • K-Folds Cross Validation

- Replaced missing values with mean for continuous variables, and general average fot categorical data.
- Took log transformation for dependent variable, price, to make it more symmetric.
- Used Ridge regression and LASSO to model the log price VS other independent variables and make a prediction.
- Implemented gradient decent algorithm and 5-Folds cross validation to tune the metaparameter and train the models.

KAGGLE: Outbrain Click Prediction
Nov 2016 - Dec 2016 | Bloomington, IN
Techincal skills: R • PostgreSQL • High Throughput Cluster • Parallel Computing • Logistics Regression

- Used PostgreSQL to join given relational tables, and extract the relevant information.
- Submitted torque script to high throughput cluster to speed up joining and extracting processes.
- Used descriptive statistics and logistics regression to model whether an ad would be clicked.

TECHNICAL SKILLS

CORE COURSES
Machine Learning • Data Mining • Advanced data structure
High dimensional data analysis • High Performance Computing
Artificial Intelligence • Bayesian Data Analysis • Time Series Analysis

PROGRAMMING LANGUAGE
R • Python • PostgreSQL • LaTeX
C(basic) • JavaSE(basic)