

Guanyan Lin

Email: guanlin@indiana.edu

Address: 1750 N Range Rd, Bloomington, IN, 47408

(willing to relocate in my own expense)

phone: 812-369-6456

Indiana University, Bloomington, IN

May 2018

MS in Data Science

GPA: 3.57

Indiana University, Bloomington, IN

May 2016

BS in Statistics, BA in Economics

Minor: Business, Mathematics

GPA: 3.74

TECHICAL SKILLS

- **Programming Language:** R, Python, SQL, JavaSE
- **Computer Application:** PostgreSQL, Microsoft Access, MySQL, Microsoft Office, PC, MAC
- **Core Courses:** Machine learning, Data mining, Advanced data structure, High dimensional data analysis, Bayesian statistics, Statistical consulting

Machine Learning and Statistics Experience

Machine Learning Research Assistant

May.2017 - Present

Daniel McDonald's lab in Indiana University 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

Statistical Consulting Assistant

Sep.2015 - May 2016

Indiana Statistical Consulting Center in Indiana University Bloomington, IN

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

Business Analytics Projects

MCD Case Presentation

Sep.2015

- Interview the owner of the manufacturer to find out what he concerned and to understand bidding, ordering, manufacturing, shipping, and paying processes
- Collect shipping logs, clock-in system, financial statements, and cost from the manufacturer and explore those data to find out how to reduce overhead cost
- Collaborated with a team of three others to provide a 45-minute presentation explaining our data based results and solution to not only layperson and professional business analyst

Kaggle Projects

Oct.2016 - Dec.2016

- Utilize R, Python, and postgresSQL to extract, transform, and clean data in different formats ranging from megabytes to gigabytes
- Implement different statistical models, such as generalized linear models, tuning the meta-parameter with cross validation criterion to find out the best model in the sense of minimizing the out of bag error based on mean squared error, if the loss function is not given
- If a particular loss function is given but I have never seen it, find the algorithms corresponding to that particular loss function in the industry literature and implement them