Zixiao Wang

Malden, MA 02148 | (571) 306 9240
wang.zixi@husky.neu.edu || linkedin.com/in/zixiaowangbrickea

SUMMARY

More than 4+ years professional studying in software engineering and data analysis. Possessing good knowledge of software development life cycle, statistical analysis and machine learning. Adept at data cleaning, data visualization and data profiling. Result oriented, quick learner seeking a business analyst role to utilize my technical and interpersonal.

SKILLS

Programming Skills:	SQL, R, Python, Java, HTML, CSS, JavaScript
Database:	MySQL
Software/Tools:	Microsoft Excel, Jupyter notebook, VSCode, Git, Markdown
Analysis model:	MLE, MOM, Bayes Estimation, T-test, Z-test, GML, SVM, Q-learning, Policy Learning

EDUCATION

Northeastern University, Boston, MA Expected Spring 2022
Master of Science in Information System (GPA: 3.556 As of now)
Relevant Courses: Application Engineering and Development, Data Science Engineering Methods and Tools, Advanced Data Science Engineering (Expected May 2020), Data Management and Database Design (Expected May 2020)

Wuhan University of Technology, Wuhan, Hubei Prov May 2019
Bachelor of science in software engineering
Honors: College Scholarship (2016), College Miyoshi Students (2016,2018)

ACADEMIC PROJECTS

Northeastern University, Boston, MA Nov 2019 – Dec 2019
NEU Skunkworks EM Lyon Workshop - *Teaching Assistance*
(Github: https://github.com/nikbearbrown/NEU_Skunkworks_EM_Lyon)

- Build the teaching notebook and help the professor to teach Students from EM Lyon to use the Jupyter notebook with models
- Using H2O to applied ML models like Deep Learning model, GLM, XRT and DRF to an ad lift dataset. Evaluate all models by comparing their RSME. Determine the important variables by calculating standardized coefficients.
- Use partial dependence plots (PDP) to show the impact of each single variable if others are kept constant. And explain the relationship base on PDP
- The final GLM model is around 0.005 of the RSME and fit the ad lift dataset very good. Finish AI workshop as one of main contributors.

Beijing PM2.5 Prediction – *Course Data Science Methods and Tools Project*. Sep 2019

- Build a random forest regression to predict future values of Beijing PM2.5 time-series dataset.
- Consider Poisson distribution as distribution of dataset base on the dependent variable probability distribution. And find the most likely parameter of pdf by using maximum likelihood estimation.
- RF can predict PM2.5 per minute within the next three hours with an accuracy of more than 75%.
- Conclude exponential distribution is more fit the data by comparing to the Poisson distribution

Wuhan University of Technology, Wuhan, Hubei Prov May 2018
"STACK" Game Development Project

- Control whole project development cycle.
- Create a 3D object cutting algorithm that is not existed in the Unity3D engine.
- Develop "STACK" video game and deploy it in the Android phone
- Gain A grade for the project