Xiaoyun Gong

xiaoyungong2022@u.northwestern.edu | https://www.linkedin.com/in/xiaoyun-gong/ | 440-412-9185

EDUCATION

Northwestern University

Master of Science in Analytics *(Data Science)* *Expected December 2022*

Expected Coursework: Databases & Information Retrieval, Predictive Analytics, Data Visualization, Data Mining, Deep Learning, Analytics Value Chain, Analytics for Big Data (Hadoop, Spark), Leadership Insights and Skills for Data Scientists, A/B Testing.

Oberlin College

Bachelor of Arts, Double Major in Mathematics (Statistics Concentration) and Visual Arts *May 2021*

Graduation with Honors in Mathematics ***GPA: 3.90/4.0, Math Major GPA: 3.95/4.0***

Relevant Coursework: Data Science and Machine Learning, Probability Theory, Bayesian Computation, Nonlinear Optimization, Statistical Modeling, Data Structures, Econometrics, Real Analysis, Group Theory, Econometrics.

PROFESSION EXPERIENCES

Duck Creek Technologies *Chicago, IL (Remote)*

Graduate Student Data Science Consultant *Oct 2021 - Current*

- Collect geographical and rating information of 20+ companies from websites and applications (Google, Facebook, etc.) via web scrapping APIs using Python; generate a Python pipeline to automize this process.
- Collaborate with SDEs on building an auto-quote platform for small businesses that increase quotation efficiency by 20%.
- Optimize the quotation platform by updating the user interface iteratively and testing new features via A/B testing.

Illuminate Health Inc. *Carmel, IN (Remote)*

Graduate Student Data Science Consultant *Sep 2021 - Current*

- Performed exploratory data analysis and extracted key metrics from over 4 million rows of medicare data using SQL.
- Built machine learning models (XGBoost, Random Forest, Lasso Regression, etc.) to predict patients at high risk with 77% accuracy rate using Python.
- Developed and deployed machine learning pipelines together with the SDE team; automated the deployment process and owned the pipeline maintainable.

DiDi Global Inc. *Beijing, China*

Data Analyst Intern *Jul 2018 - Sep 2018*

- Conducted clustering analysis model using R to segment customers into high-value, medium-value, and low-value groups; visualized each group's retention rate using Tableau Dashboard for decision making processes.
- Initiated a new coupon redemption product by collaboration with the E-commerce marketing team and SDE team.
- Designed A/B testing and A/A testing experiments to evaluate the new coupon product's influence on customer behavior increased medium-value user groups' retention rate by 16.8% in the next month.

ACADEMIC PROJECTS & RESEARCHES

Honors Thesis: Collaborative Game Mosaics Nov 2019 - Current

- An optimization art project that presents the best solution for reproducing images with different shades of gray using the rules of various chess games as constraints of the optimization problems. (Link to a virtual art show with the final product of this research.)
- Applied nonlinear optimization techniques to form objective functions and constraints and programmed a solver using C++ and Java.
- Resulted in an in-review publication on Bridges Banff: mathematical connections in art, music, and science.

Las Vegas Restaurants Inspection Project Jul 2021 - Aug 2021

- Performed data cleaning and outlier detection on 15000+ Las Vegas restaurant food establishment inspection reports using Python.
- Conducted machine learning models (KNN, XGBoost) to predict the result of next inspection at an accuracy rate of 83%.
- Created an inspection analysis report and provided suggestions to potential stakeholders.

Twitter Trend Analysis Project Jun 2020 - Sep 2020

- Created a twitter dataset through web scrapping using Python and cleaned the dataset for data integrity.
- Analyzed the dataset, displayed wordcloud and determined the top trending hashtags via topic modeling using R.
- Generated predictive trends report and put forward business development suggestions based on trends analysis.

SKILLS

Programming: Python (pandas, Seaborn, PySpark, sklearn), SQL, R (dplyr, ggplot2), Java, C++.

Data Science and Analytics: Machine Learning, A/B Testing, Time Series, Text Mining, Web Scrapping.

Cloud Computing: AWS, Microsoft Azure.