

# Haiwang Ge

hge105@syr.edu | 716-986-5065 | Syracuse, NY | [www.linkedin.com/in/haiwangge](http://www.linkedin.com/in/haiwangge)

## SUMMARY

Enthusiastic and talented individual with strong data science, and Statistics background seeking the position of Data Analyst. Bringing outstanding project and research experience in end-to-end analyzing, modeling, and interpreting data to help company and clients achieve goals and exploit opportunities.

## EDUCATION

<b>Syracuse University, Syracuse, NY</b> <i>M.S, Applied Data Science</i>	<b>Aug 2021 - Dec 2022(Expected)</b>
<b>The State University of New York at Buffalo, NY</b> <i>M.S, Material Design and Innovation</i>	<b>Aug 2019 - Dec 2020</b>
<b>The State University of New York at Buffalo, NY</b> <i>B.A.S, Mathematics, Minor: Economics and Statistics</i>	<b>Sep 2014 - May 2019</b>

## SKILL

- Analysis: Advanced SQL, Excel (lookup, correlation, regression, forecasting, 3D maps, Solver optimization) Google Analytics and AdWords, Tableau(dashboards), Power BI(dashboards)
- Data Science: Python (Pandas, NumPy, Sklearn, Matplotlib), R (ggplot2, plotly), Minitab
- Statistics: A/B Test, Hypothesis test, Z test, T test, F test, Chi-Square, ANOVA analysis
- Machine learning models: Logistical Regression, Random Forest, K-means, SVM, Decision Trees, Confusion Matrix
- Database: MySQL, Azure Data Studio

## WORKING EXPERIENCE

<b>Data Science Intern</b> <b>Built Environment Science and Technology (BEST) Lab</b>	<b>Jun 2022 – Current</b> <b>Syracuse, NY</b>
<ul style="list-style-type: none"><li>• Extracting and cleaning 218 buildings data from API and develop data-driven predictive control (DDPC) for building HVAC systems.</li><li>• Assisted to build deep neural network by Python within PyTorch framework for energy control optimization.</li><li>• Achieved the heating and cooling load saving of rooftop units system (RTUs) and variable air volume system (VAVs) by 75% and 31% respectively with above data-driven control as preliminary results.</li></ul>	
<b>Database Manager Intern</b> <b>Resilience. Inc – Social Emotional Learning Programs</b>	<b>July 2021 – Jan 2022</b> <b>Tampa, FL</b>
<ul style="list-style-type: none"><li>• Discussing project with the team during department weekly meeting</li><li>• Creating database and data mapping like ERD and EER</li><li>• Connect other department to help staff easily retrieve the information they need</li></ul>	
<b>Data Analyst Intern</b> <b>Sargent &amp; Lundy LLC</b>	<b>Jun 2017 – Aug 2017</b> <b>Chicago, IL</b>
<ul style="list-style-type: none"><li>• Performed clients segmentation analysis (RFM, K-means) to group clients into 5 types and improved design process by 25%</li><li>• Wrote SQL query in MySQL database to analyze the demand of electrical parts in multiple regions in US</li><li>• Worked with stakeholders to delivery and visualized daily, weekly, monthly reports by using Python, PowerPoint</li><li>• Analyzed sales conversion funnel to improve the contract signed rate by 5%</li></ul>	
<b>Data Analyst Intern</b> <b>Harbin Electric Corporation</b>	<b>Dec 2016 - Jan 2017</b> <b>Harbin, China</b>
<ul style="list-style-type: none"><li>• Built a time series model to predict the trend of multiple electrical parts sale in the next 6 months with 27% RMSE</li><li>• Worked with HR team to predict employee turnover rate by using logistic regression and provided insights</li><li>• Wrote an automated python code to build a people analysis report on a weekly, monthly basis and reducing preparation time by 35%</li></ul>	

## PROJECT

<b>Informatics Driven Approach for Property Prediction (Regression)</b>	<b>Jun 2020</b>
<ul style="list-style-type: none"><li>• Applied Pearson correlation and feature selection methods to analyze potential important features for predicting Bulk module</li><li>• Performed regression modeling with regression, multivariate regression, and principal component analysis to fit modes explaining the most influential factors that affect DOS spectra</li><li>• Processed multicollinearity analysis, R square, RMSE and residual plots to compare different model results in order to achieve a final prediction model with 74% <math>R^2</math></li></ul>	
<b>Crystal Structure Prediction (Classification)</b>	<b>Dec 2019</b>
<ul style="list-style-type: none"><li>• Performed EDA and data cleaning on dataset that consists of 7K+ material property data form Material Project API with labels by using python</li><li>• Conducted 3 methods to impute the missing value to achieve a best prediction performance</li><li>• Constructed five machine learning models including KNN, LOG, OVO, OVR, TREE to predict the phase diagram and crystal type, and KNN (k = 5) achieved the best 88.8% accuracy rate for phase diagram and 37% accuracy rate for crystal structure.</li><li>• Used an ensemble method to refine the existing model and improved the accuracy rate to 92.7% for phase diagram and 58.8% for crystal structure.</li></ul>	