

STACY LEE

(917) 617 9342 ☎
stacy-lee.github.io 🌐
linkedin.com/in/lee-stacy in
slee422@illinois.edu ✉

Education	University of Illinois at Urbana-Champaign M.S. in Statistics <i>Analytics Concentration</i> B.S. in Civil Engineering <i>Systems Engineering Concentration</i> <i>Minor in Mathematical Statistics</i>	Dec 2018 May 2017
Languages	Proficient: Python, R, SQL Basic: HTML5/CSS/JavaScript, SAS, MATLAB, C++	
Libraries	Python: NumPy, Pandas, Sklearn, StatsModels, Matplotlib R: Tidyverse, Caret	
Tools	AWS EC2, Redshift, Azure Data Lake, Hive, Spark, GCP, UNIX/Bash, Tableau, Git	
Experience	Ameren Data Science Innovation Intern <i>Analytics Team</i> ○ Led a 5-person team on the project for identifying individuals out of 1.4 million customers with the highest likelihood of enrolling & saving in two different energy savings programs administered by the business partner using demographic dataset of mixed data types - Built tools for exploratory data analysis to gain insights & formulate a data-driven strategy - Implemented random forest for 1% imbalanced class ratio and improved recall by 90% - Applied Bayesian statistics with logistic regression for targeted marketing deliverable - Surfaced and presented insights on customer trends through storytelling to stakeholders ○ Researched methods to replace the traditional utility pole health assessment process - Converted & reformatted raw data collected from sensors to facilitate analytics - Performed clustering methods to find patterns in pole health based on movement ○ Assisted with recommender system to estimate solar energy adoption for distribution plans ○ Executed & documented ETL processes & data pipelines on Azure or AWS EC2 for efficiency ○ Mentored teammates on machine learning, statistics, and programming in Python or R RailTEC, University of Illinois at Urbana-Champaign Research Assistant <i>Train Risk Analytics Group</i> ○ Conducted a quantitative spatial analysis on the effect of speed on derailment severity ○ Applied regression methods and developed data visuals in R for investigating trends ○ Acquired open datasets to create US kernel density maps identifying high areas of casualties ○ Debugged and implemented data integration for four datasets (over 3 million) in SQL server	Champaign, IL Jan 2018 - Dec 2018 Urbana, IL Jan 2016 - May 2017
Competitions	Synchrony Financial Datathon Top 5 of 25 Teams with Best MSE Score ○ Researched the impact of federal interest rates on home improvement spending ○ Acquired external data to combine with the given time series of expenditures ○ Implemented Elastic Net with Five-Fold Cross Validation for the final submission	April 2018
Projects	Machine Learning & Computational Statistics Projects [stacy-lee.github.io/ds/projects.html] ○ Time series with fourier seasonality on Walmart sales data achieved WMAE less than 1630 ○ Logistic regression for movie review analysis with NLP achieved AUC greater than 0.95 ○ Monte Carlo method to identify Weibull PDF in Traffic Volume Counts 2012 NYC Open Data Algorithms (Implemented From Scratch) Random Forest, k-NN, Lasso Regression, Apriori	
Involvement	Civil Engineering Undergraduate Advisory Board President ○ Established the first undergraduate advisory board in the civil engineering department ○ Organized social events of ~50 attendees to increase student and professor interactions	Champaign, IL 2016 - 2017