

STACY LEE

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

Education	University of Illinois at Urbana-Champaign M.S. in Statistics GPA: 3.62/4.00 B.S. in Civil and Environmental Engineering Systems Concentration Minor in Mathematical Statistics	Dec 2018 May 2017
Languages	Proficient: Python, R, SQL Python Libraries: Pandas, Sklearn, NumPy, Re, Seaborn, Keras	Familiar: C++, C, MATLAB, HTML/CSS, AWK, SAS R Libraries: Tidyverse
Tools	AWS EC2, GCP, Azure, Spark, UNIX/Bash, Tableau, Git	
Experience	Ameren Champaign, IL Data Science Innovation Intern Analytics Team Jan 2018 - Dec 2018 <ul style="list-style-type: none">Lead a 5-person team for the project in identifying individuals out of 1.4 million customers with the highest likelihood of enrolling in two different energy savings programs administered by Elevate Energy using Experian demographic dataset with mixed data types<ul style="list-style-type: none">Performed exploratory analysis, data cleaning & feature engineering for model productionImplemented random forest for 1% imbalanced class ratio and improved recall by 90%Presented insights on customer trends through storytelling to stakeholdersMentored teammates on machine learning, statistics, and programming in Python or RAssisted with development of prediction model to classify adoption of solar energy for distribution planningBrainstormed classification methods to replace the traditional utility pole assessment process with the goal of implementing a passive and cost-effective system<ul style="list-style-type: none">Wrote Python scripts to reformat data collected from sensors to facilitate analyticsCreated different data visuals to find a clear classification threshold based on sparse data RailTEC, University of Illinois at Urbana-Champaign Champaign, IL Research Assistant Train Safety Analytics Group Jan 2016 - May 2017 <ul style="list-style-type: none">Developed a spatial visual that predicted train derailment severity based on train speedResearched and acquired open datasets to create a kernel density map in ArcGISApplied regression methods on data regarding train casualties using R programming languageDebugged and implemented data integration for four datasets (over 3 million) in SQL queries	
Projects	Software Engineering Algorithm Implementation From Scratch <ul style="list-style-type: none">CS 412: Intro to Data Warehousing and Data Mining, Fall 2017<ul style="list-style-type: none">Random Forest Algorithm in Python (Largest dataset: 8,000 train; 4,000 test)<ul style="list-style-type: none">Used gini index and majority vote. Best test accuracy 0.96. Runtime under 5 minutes.Frequent Itemset Mining Algorithm in Python<ul style="list-style-type: none">Used apriori algorithm to output outlier resilient itemsets. Runtime under 5 minutes.	
Involvement	Civil Engineering Undergraduate Advisory Board Champaign, IL President 2016 - 2017 <ul style="list-style-type: none">Established the first undergraduate advisory board in the civil engineering departmentOrganized social events of ~50 attendees to increase student and professor interactions	
Awards	Data Science: Top 10 Team with Best MSE Score - Synchrony Financial Datathon 2018 Top 20% (Round 1); Top 50% (Round 2) - TEXATA 2017 Energy: Finalist Team - U.S. Dept. of Energy's Race-To-Zero Sustainable Design 2015 Entrepreneurial: Semi-Finalist Team - Cozad New Venture Competition 2014	