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 *Analytics Concentration, GPA: 3.65/4.00* — Expected: Dec 2018
B.S. in Civil and Environmental Engineering *Systems Concentration* — May 2017
Minor in Mathematical Statistics

Languages

Proficient: **Python, R, SQL** — Familiar: **SAS, C++, C, MATLAB, AWK, HTML/CSS**
Python Libraries: **Pandas, Scikit-Learn, Numpy, Re** — R Libraries: **Tidyverse**

Tools

AWS EC2, Spark, UNIX/Bash, HiveQL, Tableau, Hadoop, Git, ArcGIS

Experience

Ameren — Champaign, IL
Data Science Innovation Intern *Analytics Team* — Jan 2018 - Present

- 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
  - Perform exploratory analysis, data cleaning, and feature engineering for model production
  - Implement random forest for 1% imbalanced class ratio and improve recall by 90%
  - Present insights on customer trends through storytelling to stakeholders
  - Mentor teammates on machine learning, statistics, and programming in Python or R
- Develop prediction model for classifying adoption of solar energy for distribution planning
- Brainstorm classification methods to replace the traditional utility pole assessment process with the goal of implementing a passive and cost-effective system
  - Write Python scripts to reformat data collected from sensors to facilitate analytics
  - Create 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

- Developed a spatial visual that predicted train derailment severity based on train speed
- Researched and acquired open datasets to create a kernel density map in ArcGIS
- Applied regression methods on data regarding train casualties using R programming language
- Implemented data integration for four different databases (over 3 million) using SQL server

Projects

Software Engineering Algorithm Implementation From Scratch

- CS 412: Intro to Data Warehousing and Data Mining, *Fall 2017*
  - Random Forest Algorithm in Python (Largest dataset: 8,000 train; 4,000 test)
    - Used gini index and majority vote. Best test accuracy 0.96. Runtime under 5 minutes.
  - Frequent Itemset Mining Algorithm in Python
    - Used apriori algorithm to output outlier resilient itemsets. Runtime under 5 minutes.

Involvement

Civil Engineering Undergraduate Advisory Board — Champaign, IL
President — 2016 - 2017

- Established the first undergraduate advisory board in the civil engineering department
- Organized 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