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: 3.62/4.00* Dec 2018
B.S. in Civil Engineering *Systems Engineering Concentration* May 2017
Minor in Mathematical Statistics

Languages

Proficient: **Python, R, SQL** Familiar: **SAS, C++, C, MATLAB, HTML/CSS, AWK**

Python Libraries: **Numpy, Pandas, Sklearn, Matplotlib, PySpark, Re** R Libraries: **Tidyverse**

Tools

AWS EC2, GCP, Azure, UNIX/Bash, Tableau, Git

Experience

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

- Led a 5-person team on the project for identifying individuals out of 1.4 million customers with the highest likelihood of enrolling and 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
- Assisted with model development to classify adoption of solar energy for distribution plans
  - Executed regression models to investigate relations between data and labels
- Researched methods to replace the traditional utility pole health assessment process
  - Converted and reformatted raw data collected from sensors to facilitate analytics
  - Created different data visuals to find a clear classification threshold based on sparse data
- Mentored teammates on machine learning, statistics, and programming in Python or R

RailTEC, University of Illinois at Urbana-Champaign Champaign, IL
Research Assistant *Train Risk Analytics Group* Jan 2016 - May 2017

- 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 heat maps to identify concentrated areas of casualties
- Debugged and implemented data integration for four datasets (over 3 million) in SQL server

Awards

Synchrony Financial Datathon Top 10 of 25 Teams with Best MSE Score April 2018

- 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

Projects

Machine Learning & Computational Statistics Projects

- 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

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