Stacy Lee

(917) 617 9342
stacy-lee.github.io
linkedin.com/in/lee-stacy
stacy.lee013@gmail.com

Languages Proficient: **Python, R, SQL** Basic: **SAS, HTML5/CSS/JavaScript, C++**

Libraries Python: **NumPy, Pandas, Sklearn, SciPy, StatsModels, Matplotlib** R: **Tidyverse, Caret**

Tools **Google BigQuery, Azure, Hive, Spark, Tableau, AWS EC2, UNIX/Bash, Git**

Experience

Carat USA New York, NY
Analyst, Data Analytics *Impact Team* Sept 2019 - Present

- Lead the research for creating cost estimates to attain significant sample sizes for generating baselines in marketing campaign planning
- Improve online ad performance over 20% through optimal frequency & creative analyses
- Collaborate with cross-functional teams to present media performance reports with insights
- Implement XGBoost & additive time series model for production volume forecast & causal impact analysis for estimating marketing impact by including external factors from research
- Optimize Python code to be reusable for WoW or daily report automation and ad-hoc requests
- Build & maintain data pipeline using Python and SQL for media performance dashboards
- Design strategy for test & control market pairing to measure marketing ROI
- Analyze web data in Adobe Analytics & Search Ads 360 for marketing attribution

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

- Led a 5-person team project for identifying individuals out of 1.4 million customers with high propensities of enrolling & saving in energy savings programs administered by the business partner using demographic dataset of mixed data types with positive feedback from partner
  - Built custom tools for exploratory data analysis to gain insights & formulate a data 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
  - Utilized clustering methods to find patterns in pole health based on movement
- Assisted with recommender system to estimate solar PV adoption for distribution planning
- Mentored teammates on machine learning, statistics, and programming in Python or R
- Executed & documented ETL processes in AWS EC2 for efficiency

Education

University of Illinois at Urbana-Champaign

M.S. in Statistics *Analytics Concentration* Dec 2018

B.S. in Civil Engineering *Systems Engineering Concentration* May 2017
Minor in Mathematical Statistics

Competitions

Synchrony Financial Datathon Top 5 of 25 Teams with Best MSE Score April 2018

- Researched the impact of federal interest rates on home improvement spending
- Implemented Elastic Net with Five-Fold Cross Validation for the final submission

Projects

Machine Learning & Computational Statistics Projects [stacy-lee.github.io/ds/projects.html]

- Monte Carlo method to identify Weibull PDF in Traffic Volume Counts 2012 NYC Open Data
- 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

Algorithms (Written From Scratch) Random Forest, k-NN, Lasso Regression, Apriori