Stacy Lee

(646) 481 8838
stacy-lee.github.io
linkedin.com/in/lee-stacy
stacy.lee013@gmail.com

Summary *Data science and engineering professional with 4+ years of experience building predictive models & automated data pipelines for dashboards and applying data science to drive decision-making.*

Languages Proficient: **Python, R, SQL** Basic: **Scala, HTML/CSS/JavaScript, C++**

Data Tools **AWS, Hive, Spark, BigQuery, Power BI, Tableau, Azure, Databricks, dbt, Airflow, Jenkins, Git**

Education

University of Illinois at Urbana-Champaign
M.S. in Statistics *Data Science & Analytics Concentration*
Relevant Courses: Machine Learning, Computational Statistics, Data Mining

B.S. in Civil & Environmental Engineering *Transportation & Systems Concentration*

Experience

Nielsen New York, NY
Senior Data Science Analyst *Nielsen Marketing Cloud Team* June 2021 - Jan 2023
- Led the model evaluation & deployment of graph product features built on terabytes of data
- Created & evaluated varying model iterations of graph product using community detection
- Built automated dashboards to monitor segmentation models & gain weekly insights
- Scoped & planned KPI metrics for monitoring anomalies and discrepancies in data pipeline
- Debugged issues in modeling pipeline that removed data leakage & reduced query runtime
- Presented & documented model evaluation methodology & results for reproducibility
- Mentored & reviewed the Python code & queries of less experienced colleagues

Dentsu New York, NY
Data Scientist, Analytics *Marketing Effectiveness Team* Sept 2019 - June 2021
- Led the strategy for creating a new composite scoring metric using principal component analysis, selecting test & control market pairs to measure advertising impact, & creating cost estimates to project sample size baselines for Fortune 100 financial & retail clients
- Built additive time series model using over 50 variables for causal impact analysis to estimate incremental lift from media campaigns with results that received positive client feedback
- Implemented feature engineering in predictive models improving accuracy over 10%
- Oversaw end-to-end research & development of XGBoost model for marketing mix modeling
- Improved online ad performance by over 20% through optimal frequency & A/B testing
- Automated ETL data pipeline from multiple sources using APIs for daily or weekly reports
- Created data visuals to investigate discrepancies & gain insights for storytelling
- Collaborated across teams to ensure proper website tags for tracking in Adobe Analytics

Ameren Champaign, IL
Data Science Innovation Intern *Innovation 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 using demographic data of 150+ features to create a target list of recipients who will receive the promotional bill insert
  - Applied Bayesian statistics with logistic regression for targeted marketing deliverables
  - Implemented random forest for 1% imbalanced class ratio & improved recall over 50%
  - Created custom tools for exploratory data analysis to gain insights & form a data strategy
- Utilized clustering methods to find patterns in pole health based on movement from sensors

Projects

Machine Learning & Computational Statistics Projects [stacy-lee.github.io/ds/projects.html]
- Logistic regression for movie review analysis with NLP achieved AUC greater than 0.95
- Monte Carlo simulation to identify Weibull PDF in Traffic Volume Counts 2012 NYC Open Data

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