Samantha Roberts

sarroberts5@gmail.com | (603) 443-0086 | Website: samantha-roberts.github.io
For recommendations and samples of work: linkedin.com/in/samantha-roberts-profile

Skills and Capabilities

- Python: Matplotlib, Numpy, Pandas, Scikit-Learn
- R: Dplyr, Ggplot, Leaflet, Lubridate, Shiny, Stringr, Tidyr
- Web Scraping, SQL, NoSQL (MongoDB), HTML, Matlab
- Classification, Clustering, Deep Learning, Dimension Reduction (PCA, tSNE, UMAP)
- Feature Engineering, Neural Networks, Regularization (Lasso, Ridge, Elastic Net)
- AWS, GitHub, QGIS
- JMP, Microsoft Office
- Tree Based Models (Random Forests, Boosted Trees, Bagging), Supervised/Unsupervised Learning, Time Series
- Leadership, Organization, Problem Solving, Team Collaboration, Time Management
- Agile/Scrum, LEAN Six Sigma Green Belt
- Data Management, Pipelines, Warehouses & Visualizations (Hadoop, MariaDB/HeidiSQL, PySpark, Power BI, Tableau)

Education

University of New Hampshire, Durham, NH
Master of Science in Analytics and Data Science **May 2019 – May 2020**

- President of Data-Driven Women (DDW): DDW strives to build a supportive community to address the gender gap and promote the success of women within data fields
- As President: Hosted LinkedIn workshop, oversaw mentor program between current students and alumni, organized students (male and female) to attend the Women in Data Science conference

Bachelor of Science in Statistics with a minor in Spanish **August 2015 – May 2019**

- Pi Mu Epsilon (National Math Honor Society) Member: Organized and hosted Exploring Math Night, an annual event that features panel of professionals in mathematics-related careers
- Treasurer and Player of the Women's Club Ice Hockey team: Duties included budget management and budget presentation to the UNH campus recreation board along with collecting member dues
- Nourish UNH member: Provide educational programming and outreach about eating well to peers, organize and host the annual national nutrition expo at UNH, teach "What's Cooking?" classes
- Buddies Without Borders member: Pairs international and domestic students to create a space for cultural exchange, conversation, and friendships

Experience

360Intel | Manchester, NH **October 2019 – May 2020**
Data Science Consultant

- Built a clear and communicative dashboard of 54,000+ surveys for clients by creating interactive visualizations in Tableau for the mystery shopper company
- Explore quantifying survey questions and their importance to the overall annual business report through extensive data cleaning and one-hot encoding of words within questions
- Utilize Natural Language Processing and variable weighting (Tokenization, Stemming, Polarity Scores using VADER) for analysis of textual responses and dashboard creation

The Music Hall | Portsmouth, NH **October 2019 – May 2020**
Data Science Consultant

- Discovered artists that are ideal (successful) candidates for this venue through building a random forest model that aides in identifying these musicians

- Used synthetic data (SMOTE) to create 10,000 additional rows of data that allowed for train and test sets to improve modeling accuracy
- Final model resulted in 24 predictors chosen through feature importance and 86% accuracy of artist success at The Music Hall

S.A.J. Dawgs | Lebanon, NH **June 2016 – September 2018**
Proprietor

- Small business owner of food cart selling hot dogs to the public (approximately 80 customers/day)
- Balance profits while controlling expenses through Excel (started with $0 profit the first year to turning ~$3,000 profit by the end of third summer)
- Marketing of products and service through social media (Facebook), newspaper and local events

Appcast, Inc | Lebanon, NH **June 2015 – August 2017**
Intern

- Projects with Salesforce and Excel: change/consolidate salesman (25+) districts
- Customer service: respond to emails, resolve issues customers were experiencing within accounts (30+/day)

Academic Work and Personal Projects

University of New Hampshire and Durham Police, Fire and EMS PPE

- Predicting the amount of personal protective equipment (PPE) needed in response to Covid-19
- Traditional machine learning techniques (k-nearest neighbors, UMAP) to cluster county and town data from all states; Data from relevant states (similar in population/demographics) to NH used
- Data extracted from the U.S. Census Bureau and the NY Times (number of Covid-19 cases and deaths by county) to be used in analysis, along with emergency call volume data
- Communicated with relevant towns' emergency departments to receive information on change in all volume and use of PPE since Coronavirus outbreak

Neural Networks (NN) for Image Classification

- Classifying 10,000+ images using a Convolutional Neural Network by using training and test sets in Python with TensorFlow and Keras libraries
- NN had convolution, pooling and dense layers with Sigmoid activation function used for binary classification
- Resulted in an accuracy score of 98.7% for training data and accuracy score of 88.6% for testing data

Amazon Customer Reviews

- Analyzing 5 million reviews by importing data through chunking process
- Utilizing Natural Language Processing (NLP) and Sentiment Analysis (Document-Term Matrices: Count Vectorizer, TF-IDF, VADER, TextBlob, Tokenization, Stemming, polarity scores)
- Clustering techniques (K-means, DBSCAN, Agglomerative), dimension-reduction techniques (UMAP, PCA) and machine learning algorithms (KNN, Random Forest, Gradient Boost, XGBoost) used to analyze the reviews and classify each review into their respective star ratings

Analytics Summer Practicum Project

- Within a team, collected, cleaned, and merged over 60 datasets from various sources for statistical analysis using Python for clustering, PCA, KNN and R for time series analysis
- Created a profile of a person who lives in the richest county in the U.S. and the poorest county in the U.S. to compare income inequality levels and the demographics of each
- Used Tableau and JMP to produce a presentation of the analytic findings

Lending Club

- Used unsupervised machine learning to analyze the financial and socioeconomic data of the Lending Club
- Utilized Python to clean, standardize, transform, and perform statistical analysis (PCA, KNN)