Samantha Roberts

sarroberts5@gmail.com · (603) 443-0086 · linkedin.com/in/samantha-roberts-profile · samantha-roberts.github.io

Core Competencies and Skills

| Programming: Python (Matplotlib, NLTK, Numpy, Pandas, PySpark, Scikit-Learn), R (Dplyr, GGplot, Leaflet, Lubridate, Shiny, Stringr, Tidyr, Web Scraping), SQL, HTML, MATLAB **| Machine Learning/Statistics:** Classification, Regression, Clustering, Decision Trees, Deep Learning, Dimension Reduction, Feature Engineering, Hypothesis Testing, Modeling, Neural Networks, Predictive Analytics, Random Forests, Statistical Analysis, Supervised/Unsupervised Learning, Time Series Analysis **| Software:** AWS, Hadoop, JMP, MariaDB/HeidiSQL, Microsoft Office, MongoDB, Power BI, QGIS, Tableau **| Capabilities:** Agile/Scrum, LEAN Six Sigma Green Belt Certification, Data Management, Data Pipelines, Strong Communication Skills, Leadership, Organization Skills, Problem Solving, Team Collaboration, Time Management

Projects

- Neural Networks (NN): Classifying 10,000+ images using a Convolutional Neural Network. Used training and test sets; convolution, pooling, and dense layers within NN; Sigmoid activation function used for binary classification resulting in an accuracy score of 98.7% for training data and accuracy score of 88.6% for testing data
- Amazon Customer Reviews: Predicting 5-star review rating using Natural Language Processing and sentiment analysis (negative, compound, positive, neutral scores). Data imported through chunking. Included Document-Term Matrices (Count Vectorizer, TF-IDF), TextBlob, VADER, clustering techniques (K-means, DBSCAN, Agglomerative) dimension-reduction techniques (UMAP, PCA) and machine learning algorithms (KNN, Random Forest, Gradient Boost, XGBoost)
- Analytics Summer Practicum Project: Analyze income inequality using Tableau, Python (classification regression) and R. Implemented Gini index. Used time-series to analyze change in income across the United States. Created a profile of a person who lives in the richest county in the US and the poorest county in the US to compare inequality levels.
- Lending Club: Analyze data on Lending Club's financial scandal using Python (Numpy and Pandas) to clean, standardize, and perform unsupervised learning techniques (Principal Component Analysis and k-Means Clustering)
- Senior Capstone: Collect data from UNH students and analyze, using JMP Statistical Software and R, how various habits and lifestyle trends affect their GPA

Work Experience

360 Intel | Manchester, NH   October 2019 – May 2020
Data Science Consultant
- Build a clear and communicative dashboard of 54,000+ surveys for clients (Visualizations).
- Explore how to quantify survey question importance to the overall business report (Data Cleaning, One-Hot Encoding).
- Use Natural Language Processing and variable weighting (Tokenization, Stemming, Polarity Scores using VADER).

The Music Hall | Portsmouth, NH   October 2019 – May 2020
Data Science Consultant
- Identified artists that are ideal (successful) candidates for this venue.
- Used synthetic data (SMOTE to create 10,000 rows) and modeling (Linear Regression, Train/Test Sets, Random Forests, Feature Importance) to build model to aide in identifying these musicians. Model resulted in 24 predictors and 86% accuracy.

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.

StartDate Labs | 10 Water Street, Lebanon, NH 03766   June 2015 – August 2016
Intern
- Projects with Salesforce and Excel at talent acquisition company: change and consolidate salesman (25+) districts.
- Customer service: respond to emails, go into accounts to fix issues that customers are having (~30/day).

Education

University of New Hampshire, Durham, NH   May 2019 – May 2020
Master of Science in Analytics and Data Science
- President of Data Driven Women: Hosted LinkedIn workshop; oversaw mentor program; attended WIDS conference.

University of New Hampshire, Durham, NH
Bachelor of Science in Statistics with a minor in Spanish   August 2015 – May 2019
- Pi Mu Epsilon (Math Honor Society) Member: Helped organize and host Exploring Math Night.
- Treasurer and Player of the Women's Club Ice Hockey team: Budget management and presentation; collect club dues.
- Nourish UNH member: Provide educational programming and outreach about eating well to peers.
- Buddies Without Borders member: Pairs international and domestic students to create relationships.