SEAN PILI

703-835-7263 ♦ seanpili@gwu.edu ♦ Alexandria, VA ♦ https://github.com/seanpili/Main

EDUCATION

The George Washington University, Washington DC — Anticipated May 2020
M.S. Data Science (GPA 3.9)
Relevant Coursework: Machine Learning II, Cloud Computing, Natural Language Processing

Virginia Tech, Blacksburg, VA — Awarded December 2017
B.S. Statistics — **Minors:** Actuarial Science and Communication
Mu Sigma Rho Honor Society Member

TECHNOLOGY SKILLS

Languages and Libraries: Python (PySpark, TensorFlow, Keras, Pytorch, NLTK, spaCy, gensim, BeautifulSoup, pandas, numpy, sklearn, re), R (tidyverse, dplyr, gplot2, elasticnet, rvest), SQL, SAS

Tools: Linux, Git, Spark, AWS Management Console, Tableau, CQL, QlikView, Microsoft Office

RELATED WORK EXPERIENCE

Data Science Intern, *Gartner, Arlington VA* Aug 2019-Present
- Adding features to an internally created python library including NER and sentiment analysis that will allow analysts with minimal knowledge of NLP to conduct text analytics
- Increased team's internal budget by finding the driving causes of unsubscribe requests in a customer support mailbox by implementing text clustering, topic modeling, automatic cluster labeling and key phrase extraction techniques in python

Data Science Intern, *National League of Cities, Washington, DC* Mar-Aug 2019
- Translated a policy model for researchers to simulate the effects of policy changes on the U.S. population from SAS to R
- Debugged python scripts coworkers used to do data wrangling and machine learning
- Taught researchers to implement penalized regression to handle multicollinearity
- Initiated and provided strategies to complete a project that would automate the tagging of future goals in mayoral speeches that the NLC is currently implementing

Reporting Analyst Intern, *Merchant Link, Silver Spring, MD* Jun2018-Jan2019
- Performed customer segmentation to help the company better service their customer base
- Generated and improved the quality of periodic KPI reports in Excel
- Updated pricing models and conducted market research to identify new business partners

Undergraduate Research, *Virginia Tech Biocomplexity Institute, Arlington, VA,* May-Aug 2017
- Participated in their Data Science for the Public Good program and worked on three social science focused research projects, one of which entailed finding the best method(s) to create a synthetic dataset for sensitive, public policy data.
- Used R to web-scrape, query, clean, aggregate, visualize and analyze data for all projects.

Product Classification Intern *USPS, Washington, District of Columbia* May-Aug 2016
- Updated their 2013-14 First-Class Mail (FCM) Product Plan for FY 2017
- Conducted a feasibility analysis on the introduction of a Standard Mail Card Product

PROJECTS

DATS 6450 Natural Language Processing: Final Project
- Goal: identify clusters of similar characters across multiple movies and create a model that can predict whether a movie will be successful (positive ROI) or a flop (negative ROI)
- Used python to scrape the scripts of 700 films and ROI data, map each film to it's ROI, extract characters from scripts, extract features for the clustering and modeling tasks, conduct LSA and DBSCAN for clustering, and train the predictive models
- Result: DBSCAN was able to successfully identify clusters of similar characters across 700 films of different genres, but the accuracy of the final predictive model was not above the no information rate. *(Code and results can be found on the linked github repository.)*