PAIGE MCKENZIE

PaigeMcKenzie@utexas.edu • https://p-mckenzie.github.io/

EDUCATION

The University of Texas at Austin — Master of Science in Business Analytics — May 2018
Overall GPA: 3.71

The University of Texas at Austin — Bachelor of Science and Arts in Mathematics — May 2017
Business Foundations Program (Highest Distinction)
Overall GPA: 3.96; summa cum laude; Phi Beta Kappa

EXPERIENCE

NetApp – *Data Scientist;* Research Triangle Park, NC — April 2019 – present
- Using data to enable sales and marketing priorities

Cisco Systems – *Data Analyst;* Research Triangle Park, NC — July 2018 – April 2019
- Prototyped Logistic Regression to classify onion websites by content type, using features engineered from page text
- Implemented topic modelling to identify trends in customer-submitted content parsed from 4 years of RFIs
- Automated data extraction and cleaning of raw text in customer case descriptions, providing visibility to previously untracked metrics and demonstrating errors in the established manual process of data collection

The University of Texas at Austin – *Workday Analyst Intern;* Austin, TX — May 2016 – May 2017
- Wrote Python program to validate campus-provided data against Workday, saving over 80 hours of work on each data load
- Reviewed over 300 business process security policies using Python, discovering and correcting the 44% error rate
- Analyzed 42,000 security role assignments using Python, finding and eliminating 1,600 incorrect assignments

PROJECTS

3M Accounts Receivable Capstone — Spring 2018
- Analyzed accounts receivable balances and invoices for all of 3M's US-based subsidiaries in 2013 through 2017
- Identified seasonality in sales and trends in invoice terms that contribute to decreasing AR turns

Instacart Market Basket Analysis — Fall 2017
- Engineered 30 features for over 13 million user-product pairs to measure product affinity
- Implemented random forest, neural network, and XGBoost models to predict whether a product would be ordered again
- Achieved a 20% improvement in AUC over a naive forecast

Personal Blog (ongoing) — Fall 2017 - present
- Discuss personal projects related to data science, including descriptive and predictive analyses
- Document code samples for creating visualizations, conducting analyses, and data scraping
- Available at https://p-mckenzie.github.io/

TECHNICAL EXPERTISE

Languages: Python, working proficient in SQL, exposure to R, SAS
Machine learning: Multivariate Regression, Clustering (k-means, agglomerative hierarchical), Binary and Multi-class Classification (logistic regression, naive bayes, decision trees, random forests, support vector machines, simple neural networks)
Data visualization: Matploblic, bokeh, Tableau, exposure to D3.js
Cloud computing: Exposure to Amazon Web Services (MapReduce, S3), Google Cloud compute, Databricks (Apache Spark / PySpark)

COMPETITIONS

USAA Hackchat – *2nd Place* — Spring 2018
- Analyzed customer service call transcripts in team of 4 students
- Used lemmatization and tf-idf to classify callers by the supporting document the call addressed

Citadel Data Open at Texas – *3rd Place* — Spring 2018
- Analyzed NYC taxi cab/Uber data for 2014-2015 in team of 4 students
- Identified and reported on demographics in neighborhoods where Uber created demand rather than cannibalizing taxi trips