PAIGE MCKENZIE

Data Scientist

🌐 p-mckenzie.github.io

🐙 p-mckenzie

✉ paige.a.mckenzie@gmail.com

EDUCATION

MS, Business Analytics

University of Texas at Austin

📅 May 2018   🎓 3.71/4.00

- -

BSA, Mathematics

University of Texas at Austin

📅 May 2017   🎓 3.96/4.00

TECHNICAL SKILLS

Programming

General:

Python, SQL, git, Jupyter notebooks, Ruby

Data Visualization:

Matplotlib, tableau, mode

Data Analysis:

NumPy, Pandas, scikit-learn, NLTK

Big Data:

PySpark, Google Cloud Platform (GCP)

Machine Learning

Supervised:

Generalized Linear Models, Tree Based Models, Nearest Neighbors, Support Vector Machines, Simple Neural Networks

Unsupervised:

Clustering, Principal Component Analysis

EXPERIENCE

Stitch Fix

Machine Learning Data Scientist   📅 September 2022 – present

- Supported re-work of on-hand inventory forecasting capabilities

- -

Shopify

Software Developer   📅 March 2022 - September 2022

- Led technical design and implementation of pricing & custom offer experiments, including experiment group assignment, data logging, and graphQL endpoints
- Led database cleanup effort for Ruby on Rails application, including removing deprecated Ruby code while handling data-lake imports & data retention

Data Scientist   📅 September 2021 – March 2022

- Updated PySpark modules to be time-aware, improving resiliency to data outages & ensuring consistency in data delivered to external partners
- Developed & updated data pipelines, including strategies for data flow, retention, and aggregation

- -

NetApp

Data Scientist   📅 April 2019 – September 2021

- Led deployment of machine learning model to forecast system utilization for over 250K systems, replacing legacy process and improving prediction accuracy by >42%
- Worked cross-functionally with sales leadership to conduct direct-to-field sales campaigns based on machine learning recommendations
- Championed data warehouse initiative, including pushing for standardization of model inputs, clearer retention policies, and transparent model evaluation & monitoring

- -

Cisco Systems

Data Analyst   📅 July 2018 - April 2019

- Automated data extraction from >4 years of unstandardized customer-submitted files, extracting useful text and implementing topic modeling to discover trends in customer inquiries

PROJECTS

Recommending TV Shows via Collaborative Filtering   link ↗

- Built a recommendation engine using users' TV show ratings to predict their rating of new shows
- Achieved 15%-32% improvement in prediction accuracy across 3 different shows, while reducing required data size by 85%

- -

AI Learns to Play 2048   link ↗

- Programmed the sliding puzzle game 2048, including data logging and user interface
- Implemented Monte Carlo Tree Search & reinforcement learning algorithm to learn optimal strategies for playing 2048

- -

Text Generation with Markov Chains   link ↗

- Designed program to generate believable sentences in different styles
- Implemented text processing to train Markov Chains of variable size with unprocessed & unstructured input text