PAIGE MCKENZIE

Data Scientist

p-mckenzie.github.io

p-mckenzie

paige.a.mckenzie@gmail.com

paige-mckenzie

TECHNICAL SKILLS

Programming

General:
Python, SQL, Git, Jupyter Notebooks

Data Visualization:
Tableau, Bokeh, D3.js

Python Packages:
Pandas, Sklearn, Matplotlib, NumPy, NLTK, PyTorch

Big Data:
Apache Spark (PySpark), Apache Hadoop (Hive)

Machine Learning

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

Unsupervised:
Clustering, Principal Component Analysis

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

EXPERIENCE

Data Scientist
NetApp
April 2019 – present Raleigh, NC

- Leveraged Pytorch on GPU to predict customer purchase behavior, achieving a 72% reduction in model training time compared to existing process
- Engineered models to predict timing of customer purchases, achieving a 5x lift in accuracy over a random forecast
- Reworked codebase left over from a short-term consulting project, producing updated models while simplifying & documenting opaque processes

Data Analyst
Cisco Systems, Inc.
July 2018 - April 2019 Raleigh, NC

- Developed model to classify websites by content type, based on unstructured 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

PROJECTS

AI Learns to Play 2048 link

- Programmed game logic for 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
- Achieved maximum tile of 2048 in 60% of games, and next-highest tile of 1024 in the remaining 40%

Targeting At-Risk Restaurants for Better Sanitation link

- Analyzed restaurant inspection history from Wake County Open Data to identify restaurants at risk of performing poorly on their next inspection
- Simulated marketing campaign that successfully targeted at-risk restaurants 4x more often than a random sample

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%

Tracking Live Audience Reactions on Twitter link

- Analyzed tweets acquired via Twitter API to track audience reactions to plot developments in the season finale of a popular reality television series
- Performed sentiment analysis & word association to identify characters whose reputations were positively or negatively impacted