Benjamin Becze

(818)-518-7934 ● Escondido, CA ● bbecze242@gmail.com ● https://github.com/Benjamin242
Website: https://benjamin242.github.io

PROFICIENCIES

- AWS cloud computing with EC2, EMR, S3 and Azure databases
- Analysis in Python, JMP, or Excel.
- Machine Learning: decision trees, ID3 decision trees, logistic regression, perceptron algorithms, and boosting algorithms
- Deep Learning: CNNs, LSTMs, RNNs, GNNs
- Data Types: Time Series, Graph, geospatial
- Data visualization techniques
- GitHub and version control

SKILLS

PYTHON | SQL| EXCEL | PANDAS | TABLEAU | PYTORCH | ALTAIR | AZURE | DASK | R | GITHUB | JAVA | ARCGIS | WEB APIS | JMP | AWS | DOCKER

EDUCATION

UC San Diego | Data Science major | Cognitive Science Minor

- B.S. in Data Science with minor in Cognitive Science

EXPERIENCE

Data Analytics/Data Science Consulting

December 2021 - Ongoing

VSNew | Remote

- Data analysis in python of time series data queried from PostgreSQL database.
- Cleaning and shaping data, dealing with missingness, transforming data with python scripts using pandas, numpy, nltk packages.
- Extracting useful information and gaining insights from the data. Created data visualizations to convey changes and insights in time series data.
- Using web apis, and creating web scrapers in python to parse through web data for specific needs.
- Transforming audio data to machine learning ready format for deep learning audio classification with pytorch.
- Created new solutions for complex string matching/searching algorithms using natural language processing techniques.

Projects

Graph Neural Network Based Spotify Recommender | Remote — September 2021 – March 2022

- Collaboration with peers Shone Patil and Jiayun Wang to use deep learning methods on graph data to create a recommender for personalized song playlists with Spotify data.
- Queried Spotify web api for feature data on a large scale, and made data processing pipelines to properly shape data for machine learning and graph creation.
- Created GraphSAGE embeddings and multi layer perceptron classifiers in pytorch.
- Analyzed recommender results and link prediction results and created readable and informative visualizations.
- Website: https://shonepatil.github.io/GNN-Spotify-Recommender-Website/
- Report: https://drive.google.com/file/d/1AWSRZxtrkEssVRl34V5YTdRHZMGPGxVk/view?usp=sharing

UC San Diego 2020 COVID-19 Data Challenge in Border Communities |Remote — July 2020 - September 2020

- Used web data to determine risk factors for school openings in San Diego.
- Combine data from many different sources, joining on spatial features.
- Created a K-means clustering algorithm in python to cluster based on feature data.
- Created many maps and visualizations that deliver the results efficiently.
- GitHub Repository: https://github.com/renaldyh27/COVID-Cool-for-School
- Story-Board: https://storymaps.arcgis.com/stories/aaeead0a241947cda83199336118087a