

Benjamin Becze

Data Scientist

CONTACT

(818) 518-7934

bbecze242@gmail.com

bbecze.com

Linkedin:

<https://www.linkedin.com/in/benjamin-becze-307b541b6/>

EDUCATION

BS in Data Science

Cognitive Science

Minor

UC San Diego

Software

Python

Javascript

Docker

Spark

AWS

Azure

PostgreSQL

Libraries

PyTorch

Transformers

Tensorflow

VLLM

OpenAI API

Langchain

Pandas

Scikit-Learn

Flask

React

EXPERIENCE

Hindsait - Data Scientist

Oct 2022 – Present

- Developed and lead team in integration of on-prem LLM solutions into rule based text extraction engines for unstructured medical data.
- Trained models for auto and semi-automated clinical decision making for Medicare authorization requests, saving clients several thousands per year in costs.
- Improved existing medical extraction engine for medical authorization criteria, resulting in a 30% reduction in clinical review time.
- Built image classification and figure extraction for medical texts including MRI scans, handwriting, visual field diagrams etc.
- Developed full-stack solutions, building client-facing frontend interfaces to complement backend data pipelines and models.

VSNew = Data Scientist/Consultant

Dec 2021– Present

- Developed scoring/ranking algorithms based on vehicle data for public databases.
- Cleaning and shaping vehicle data and engine audio data for audio classification with PyTorch.
- Analyzing time series vehicle data for insights from various test drives.

Projects

Graph Neural Networks for Spotify Recommender

Sep 2021 - Mar 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.
- Created GraphSAGE embeddings and multi layer perceptron classifiers in pytorch.
- <https://shonepatil.github.io/GNN-Spotify-Recommender-Website/>
- <https://drive.google.com/file/d/1AWSRZxtrkEssVRL34V5YTdRHZMGPGxVk/viiew?usp=sharing>

UC San Diego 2020 COVID-19 Data Challenge in Border Communities

Jul 2020– Sep 2020

- Used web data to determine risk factors for school openings in San Diego
- Combine spatial data with public CDC data, joining on spatial features.
- Created a K-means clustering algorithm in python to cluster based on feature data.
- <https://github.com/renaldyh27/COVID-Cool-for-School>
- <https://storymaps.arcgis.com/stories/aeeadoa241947cda83199336118087a>