Jacob Cutter

San Ramon, CA | 510-861-0706 | jecutter92@gmail.com | LinkedIn | **Website** | **GitHub**

SKILLS

Programming/Databases: Python, C++, BASH/UNIX shell, SQL (MySQL, PostgreSQL, SQLite), Github
Techniques (Tools): Machine Learning (Scikit-Learn, PyTorch), ETL (SQL, Pandas, NumPy), Visualization (Matplotlib, Seaborn, Jupyter, Bokeh), Apps (Flask, Docker), NLP (NLTK, spaCy), Statistical Modeling and Analysis, Signal Processing, Data Processing/Management (Grid Computing, SGE, SLURM, RAID)

EXPERIENCE

Deepgram, San Francisco Bay Area, CA

Data Scientist, Team Lead of Product Development Oct 2022 - Present

- Coordinate cross-functional efforts across Product, DataOps, Research and Engineering departments to optimize AI model R&D and expand our ASR languages and product offerings to customers
- Employ the team's latest AI architectures and tooling to train and productionalize new ASR models
- Maintain IC workload while managing a team of researchers to execute projects that have a direct impact on the revenue stream through customer acquisition and upsell opportunities

Data Scientist Oct 2020 - Oct 2022

- Trained and maintained dozens of iterations of E2E ASR models using PyTorch frameworks
- Devised and reported KPIs to assess our core speech products, and internally evaluated novel next-generation deep learning architectures to optimize ASR accuracy/performance
- Spearheaded efforts to improve ASR punctuation by building out multilingual pipelines for text cleaning, data preparation, and experimental ML modeling with PyTorch
- Sourced, curated and preprocessed TBs (1000s of hours) of unstructured speech data for research
- Mined production PostgreSQL databases and leveraged customer metadata to construct representative training/evaluation datasets and facilitate targeted product R&D
- Assisted with the development of Flask applications for NLU products such as text summarization

Insight Data Science, *Data Science Fellow,* San Francisco CA May 2020 - July 2020

- Created a music classification app for listeners and content creators to filter Spotify playlists by emotion
- Crowdsourced emotional labels from Last.fm SQLite databases and combined Spotify audio features with song lyric sentiment to build song emotion classifiers with up to 71% accuracy
- Deployed the classification models on AWS in an interactive Streamlit web application

UC Davis Physics Department, *Graduate Student Researcher and T.A.* Sept 2014 - August 2020

- Designed local R&D particle experiments to statistically model important nuclear processes
- Developed end-to-end C++/Python pipelines for signal processing, data reduction and visualization, and synthesized TBs of unstructured, noisy waveform data into physical measurements
- Used MySQL database replication and Flask to remotely monitor lab devices via web interface

PROJECTS

NBA Basketball Analytics | github.com/jecutter/nba-data-models, jecutter.github.io/blog/ Jan 2020

- Web-scraped many seasons of NBA player, lineup, and play-by-play data using Scrapy and Selenium
- Created an interactive visualization dashboard using Bokeh to explore player and lineup data
- Built a Random Forest classifier model to perform player comparisons with 92% accuracy
- Used ridge regression on harvested PBP data to develop a lineup-independent player impact metric

EDUCATION

University of California, Davis August 2020
Ph.D. in Experimental Particle Physics, Designated Emphasis in Nuclear Science

University of California, Davis June 2014
B.S. in Particle Physics