Jacob Cutter

Mountain View, CA | 510-861-0706 | jecutter92@gmail.com
linkedin.com/in/jacob-cutter/ | **jecutter.github.io/** | **github.com/jecutter**

SKILLS

Programming/Databases: Python, C++, BASH/shell, SQL (MySQL, PostgreSQL, SQLite), Github/GitLab
Techniques (Tools): Machine Learning (Scikit-Learn, PyTorch, Keras), ETL (SQL, Pandas, NumPy), Visualization (Matplotlib, Seaborn, Bokeh, Flask), NLP (NLTK, spaCy), Statistical Modeling and Analysis (SciPy, Statsmodels), Signal Processing, Data Pipelining and Management (Computing Clusters, RAID)

EXPERIENCE

Data Scientist, *Deepgram,* Mountain View CA — Oct 2020 - Present

- Training and monitoring state-of-the-art deep learning ASR models using PyTorch frameworks
- Reporting KPIs for core ASR models and devising evaluation metrics to track performance
- Developing and optimizing parallelized Python tools for model training, testing, and data analysis
- Wrangling, cleaning, and pre-processing speech data for model training and evaluation
- Curating TBs of audio data, including automating Q.A. and classification processes
- Syncing with Data Ops, Product, and Engineering teams to coordinate cross-functional efforts for optimizing ASR model R&D and standardize data flow for the Research Team

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

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

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

- Spearheaded and designed local R&D particle experiments, using statistical analysis and modeling to characterize important nuclear processes and inform major dark matter searches
- Developed custom end-to-end C++ and Python software for signal processing, data reduction and visualization, synthesizing TBs of noisy waveform data into physical measurements
- Used MySQL database replication and Flask to remotely monitor lab operations via web interface
- Managed large-scale data processing pipelines for an international multi-million dollar experiment
- Devised and taught lesson plans for discussion/lab sections for multiple undergraduate physics courses, receiving excellent student evaluations averaging higher than 4.5/5 across the board
- Mentored ~75 high school students over 3 summers as part of the COSMOS program, working with students to execute intensive data-driven astrophysics projects in just 1 month

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 useful player comparisons with 92% accuracy
- Developed an RAPM model using ridge regression on matchup results harvested from play-by-play data, providing a lineup-independent impact metric for scouting undervalued players

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