Adam P. Jones, PhD

www.adam-p-jones.com

ajones173@gmail.com
adam-p-jones
meccaLeccaHi

Experienced data scientist proficient in the interpretation and visualization of data using Python and R. Proven ability to work either independently or as part of a team, and to communicate results in a precise, intuitive format to stakeholders of various technical backgrounds. Intent on applying these skills to data science problems, particularly those that seek to benefit to society.

Experience

5/2020-Present | Data Scientist | Intuit

- Analyze effects of current marketing strategy via mixed media modeling and time series forecasting, in order to better inform brand building efforts across the organization.
- Advise Brand Analytics team on developing more robust data science practices.

3/2018-Present | Lead Instructor | General Assembly

- Distinguished Faculty Member for Data Science and Python training programs.
- Developed course content on a variety of technical subjects and mentored students (from companies including Intuit, eBay, and LinkedIn) through individualized projects on topics such as time-series forecasting, recommendation systems, fraud detection, and NLP.

6/2017-6/2018 | Lead Data Scientist | Critical Juncture

- Identified strategies, via academic literature review, for improving the accuracy of a medical record linkage system providing clinical performance metrics to more than 200 hospitals.
- Implemented probabilistic record linkage algorith to match records from multiple databases, reducing non-matched records by $\approx$75%. Presented results via invited lecture.
- Trained convolutional neural network to locate and classify images embedded within the digital archives of the Federal Register (with $>$98% accuracy), improving the readability and accessibility of decades of government documents.

1/2016-5/2017 | Post-doctoral Researcher | Neurosurgery - U. of Iowa

- Implemented image/sound processing pipeline for realistic "morphing" between different faces and voices (using Python and OpenCV), for presentation in neurophysiological studies.
- Developed 'gamified' data collection platform, which integrated joystick and eye-tracker input, resulting in $\approx$50% greater participation by pediatric patients.
- Deployed browser-based surveys via AWS, reducing data collection costs dramatically.

10/2012-10/2015 | Pre-doctoral Research Fellow | National Institutes of Health

- Designed, deployed, and maintained a data processing pipeline for large volumes of electrophysiological data, which included dimensionality-reduction and clustering of neural events.
- Used ML to decode the identity of faces from time-series data (the responses of visual neurons). Observed changes in decoding accuracy throughout intensive behavioral training.
- Presented results via invited lectures (3), posters (8), and written reports (4 journal articles).

Skills

Python (NumPy, pandas, Keras), R (dplyr, ggplot2, Rmarkdown), SQL, Jupyter, UNIX, Flask, LaTeX, Matlab, parallel processing (TensorFlow/Theano), distributed computing (cluster, AWS), hypothesis testing, Bayesian statistics, machine learning, neural networks, image processing, signal processing (spectral analysis)

Education

9/2009-12/2015	PhD (Neuroscience)	University of Maryland
9/2002-4/2007	BA (Biology/Psychology)	University of Montana

- Coursework included statistics, calculus, linear algebra, quant. analysis, signal processing

Recent Projects

- **Neural networks – Cyborg moths** (Summer 2019)**:** This project involved a biologically-inspired neural network, which simulates the architecture and temporal dynamics of the insect olfactory system. It does so by running a time-stepped simulation of learning in the moth brain, in which it is trained on a cruder, non-spatial version of MNIST. The model is meant to match with the architecture of the natural moth brain, in terms of size and complexity ($\sim$60 features).

  This package (*Pymoth*) compares the predictions of the moth brain to those of kNN and SVM models. We find that our moth brain *out-performs the other ML methods when trained on $<$20 training samples* per class (*link to article*).

- **Record linkage – DataKind/Muso Health** (Fall 2019)**:** I designed a semi-supervised approach to the linkage of survey records spanning three years of a controlled trial testing the impact of child mortality prevention efforts in Mali.

  This was made particularly challenging due to the increasing instability of the region. Since, due to migration and marriage patterns, study participants frequently used multiple IDs. The lack of precise training labels also required extra creativity when assessing model performance.

- **Image classification – Federal Register** (Spring 2018)**:** In order to improve the overall readability and discoverability of decades of government documents, I trained two image classification models: one to discriminate semantic category and the other the rotation of figures embedded in the digital archives of the Federal Register.

  They each consisted of a pre-trained convolutional neural network with drop-out, and a partially un-frozen convolutional base, both of which *perform with very high ($>$98%) accuracy*.

- **Record linkage – CA Maternal Data Center** (Fall 2017)**:** As part of the creation of a statewide perinatal records database, aimed at improving the quality of maternal healthcare, I designed a linkage algorithm that made use of numerous probabilistic/ML algorithms to match de-identified records from the same individual.

  After implementation, there was a *$\approx$75% reduction in non-matched records*, providing more accurate clinical performance metrics to more than 200 hospitals. Presented results via invited lecture at PyBay 2020.