REBECCA KRALL, PH.D. DATA SCIENTIST

rebecca.krall@gmail.com
407-449-3010
Mountain View, CA
rebeccakrall
rebeccak1

EDUCATION

Harvard University
Ph.D. Theoretical Particle Physics 2017

University of Cambridge
Master of Advanced Study Physics 2013

Carnegie Mellon University
B.S. Physics 2011

SKILLS

PROGRAMMING LANGUAGES & SOFTWARE
Git
Python
Shell
SQL
Unix

MACHINE LEARNING & STATISTICS
linear/polynomial/logistic regression
XGBoost
k-NN
random forests
parametric tests

TOOLS
Jupyter
matplotlib
numpy
pandas
seaborn
scikit-learn
scipy

AWARDS

NSF Graduate Research Fellowship
National award for 3 years of graduate study (declined)

DOE Office of Science Graduate Research Fellowship
National award for 3 years of graduate study

Churchill Scholarship
National award to 15 STEM college seniors for one year at University of Cambridge

EMPLOYMENT

Facebook — Menlo Park, CA
Research Data Scientist — July 2019 to Current

- Create and deploy big data pipelines to transform mobile-collected data and models to predict users' home ISP and download speed (Presto, Python, SQL)
- Interpret results from A/B tests and recommend next steps to stakeholders
- Develop new metrics for measuring mobile and WiFi network quality

Continental Finance Company — Wilmington, DE
Data Scientist — Apr. 2018 to July 2019

- Trained a model to predict credit default and give a credit limit increase without directly linking customers to their personal credit information, improving KS of model by 10.
- Developed a credit card offer response score for customers to help prioritize marketing
- Built FCRA compliant credit default risk prediction models (scorecards) for consumer new account underwriting decisions (logistic regression, Python, random forests, SQL, xgboost)
- Collaborated with IT team to implement models

Insight Data Science — Boston, MA
Data Scientist Fellow — Sept. 2017 to Dec. 2017

- Worked as a consultant for the company AnimalBiome to predict sick/healthy cats with microbiome data
- Applied dimensionality reduction with PCA to reduce the number of features to less than the number of samples, and used logistic regression on reduced feature set (Git, Python, numpy, pandas, scikit-learn, scipy, SQL)

Legendary Entertainment — Boston, MA
Quantitative Research Intern — Mar. 2016 to May 2017

- Completed two projects: age and gender prediction, and logo detection and classification
- Trained a neural network (OpenFace) using 400,000 images to embed faces on a 128-dimensional unit hypersphere
- Predicted gender with SVC on neural net features, achieving 90% accuracy
- Improved age prediction F1 score by 30%, compared to the machine learning methods in use, using multiple linear regression on neural net features (Lua, OpenCV, Python, numpy, pandas, scikit-learn, scipy, Torch)
- Trained production-ready Faster R-CNN for logo detection and classification

Wolfram Research Inc. — Somerville, MA
Advanced Research Group Intern — July 2015 to Dec. 2015

- Developed a LSTM neural network for speech recognition (C, Git, Lua, Mathematica, Python, Torch)
- Implemented Connectionist Temporal Classification loss function (used in best speech recognition system at time)

Harvard University — Cambridge, MA
Ph.D. Research Assistant — June 2013 to May 2017

- Wrote code to simulate particle decays and predict experimental results (C++, Python, Shell)
- Modified open source cosmological simulation code of 300,000 lines to include new particles in cosmological model
- Applied C and Python implementation of Markov Chain Monte Carlo (MontePython) to determine the best-fit parameters of a cosmology model