


Aaron Gonzales

agonzales@cs.unm.edu | 505.385.9209 | aarongonzales.net | in | 

About

Results-oriented data scientist with experience building scalable machine-learning systems. Comfortable working with large data sets (billions and billions of observations; tera-to-petascale data) and practicing end-to-end R&D (design, prototype, test, deploy, improve).

Selected Experience

Data Scientist, TripAdvisor

July 2016–Current

- Built an internal app that supports natural-language queries (e.g. *"family-friendly beaches"*) to find collections of destinations using current NLP word embedding techniques.
 - ◊ Used by Content Marketing to assist with content creation.
 - ◊ Used by SEO to improve relevance (A/B tested with **4.8% improvement over control in a primary metric**).
 - ◊ Flights team is testing it to recommend nearby destinations for travelers.
- Designed and built a computer vision pipeline for all photos on the Rentals platform using deep convolutional networks and feature learning, with flexible methods to share learned data across teams.
 - ◊ Developed model to flexibly detect people in photos with and without faces (91% test accuracy).
 - ◊ Developed scene and content classification model for all photos (94% test accuracy).
 - ◊ Developed image aesthetic quality assessment model (84% test accuracy).
 - ◊ Developed model to understand how photo content affects traveler behavior.
 - ◊ Display marketing saw an A/B-tested **4.8% increase in click-through rate** and a **10.4% increase in traveler inquires**.
 - ◊ Data delivered by the pipeline is in use across the Rentals platform.
 - ◊ Models will likely be utilized across other key TripAdvisor business areas, including Content and Hotels.

Data Scientist Intern, TripAdvisor

Summer 2015

- Developed a novel ranking system for 720,000+ rental properties using creative feature generation and gradient-boosting machines, which went into production after successful A/B testing.
- Model was built using approximately 5 terabytes of traffic history using Hive, Python, pandas, and scikit-learn.
- Model scales with millions of daily visitors and self-tunes to fluctuations in visitor usage patterns.
- A/B testing showed that the model **increased a key visitor conversion rate by 3.46%**, **decreased visitor bounce rate by 0.46%**, and **increased revenue per visitor by 9.57%**.

Analyst/Programmer, The University of New Mexico

2011–2014

- Streamlined lab data processing and analytical techniques, including a method that improved execution time for a critical data processing step by approximately 360x (3 hours to 30 seconds). Implemented numerous other batch data processing steps for other tasks.
- Designed and conducted a pilot study that helped win a \$2.7 million dollar NIH R01 grant to study the etiology of post-traumatic stress disorder.
- Contributor and author on several research papers, conference abstracts and presentations. Publications are available on request, LinkedIn, or my site.
- Trained and mentored 10 undergraduate and post-baccalaureate student employees and volunteers

Education

Master of Science, Computer Science, The University of New Mexico

2016

Bachelor of Science, Psychology, The University of New Mexico

2010

Technical Skills

Programming languages, notable libraries, and tools

R, Python (Numpy, Scipy, scikit-learn, TensorFlow, gensim, Pandas, Statsmodels, Cython, SQLAlchemy, Keras, Jupyter, Bokeh, seaborn), Spark, Pyspark, C, Bash, \LaTeX , git, MongoDB, Hadoop, Hive, SQL (Postgres, Sql Server), Linux (Ubuntu/Centos) admin, cloud/distributed system development. Some experience with Scala, Java, Javascript, Matlab, and Amazon Web Services (EC2, S3, Redshift).