# David Bang

Data Scientist

## Education

**2018-08 - 2020-05** **University of North Carolina at Chapel Hill**
- M.S. / Ph.D. in Biostatistics

**2012-08 - 2016-05** **Vassar College**
- B.A. in Economics

## Experience

**2018-05 - 2019-01** **Analytics Intern**
*Valassis Digital*
- Developed a data visualization dashboard to display ad performance using Plotly & Dash hosted on a Flask server
- Created ETL pipelines to help automate common tasks
- Employed Multi-Class Logistic Regression to predict bidding price settings for line items with performance metrics as features
- Won company hackathon against 15+ teams by implementing a custom clustering algorithm on line items to identify similar ad groups
- **Received return offer for the fall semester**

**2018-07 - 2018-08** **Bioinformatics Research**
*Duke University*
- Selected to be a part of a 4 person NIH-funded research program in computational biology and bioinformatics.
- Employed DESeq & gene set analysis for *Cryptococcus neoformans* RNA-Seq.
- Supervised Learning: logit, spline reg, cross-validation
- Unsupervised Learning: MDS, PCA, AHS, k-means, noise discovery

**2018-05 - 2018-08** **Mentored Research**
*UNC Chapel Hill*
- Economics Research

**2016-01 - 2016-05** **Mentored Research**
*Vassar College*
- Economics Research

## Projects

**2019-01** **Twitter Sentiment Analysis** | Natural Language Processing
**Tech Stack:** Python, MySQL
- Streamed tweets using Twitters API via tweepy and stored it in a MySQL database
- Employed Naive Bayes Classifier to determine positive/negative sentiment
- Cleaning: stop words, speech tagging, chinking, chunking, etc.

**2018-12** **Movie Lens (1M)** | Recommendation Systems
**Tech Stack:** Python, Tensorflow, Keras
- Content-Based Filtering
- Memory-Based Recommendation System: Matrix Factorization
- Model-Based: User-User Collaborative Filtering
- Model-Based: Item-Item Collaborative Filtering

**2018-10** **Potato Classifier** | Computer Vision
**Tech Stack:** Python, Tensorflow, Keras, Google Cloud
- Downloaded 5000+ original photos from Google Image Search and ImageNet of potatoes and not potatoes for training, test, & validation data. Trained it on 30,000+ augmented photos.
- Implemented a Stochastic Gradient Descent model on the Google Cloud for potato classification. Ended up with 98% accuracy for internet photo prediction but 35% accuracy for live photo prediction. (work in progress)

## Hackathons and Tech Conferences

**2018-10** All Things Open

**2018-10** HackNC

**2018-03** HackathonCLT

## Contact


**Phone**
919-633-4148

**E-mail**
handavidbang@gmail.com

**Website**
www.quirq.io

**GitHub**
github.com/handavidbang

**LinkedIn**
linkedin.com/in/handavidbang


## Languages

Python, R, SQL

## Technologies

Spark, Hive, Docker, Google Cloud Platform, TensorFlow, Keras

## Skills

Linear Regression & ANOVA

(Multi-Class & Binary) Logistic Regression

Regularization: Ridge & Lasso

Principal Component Analysis

Least Angle Regression, Elastic Net

Decision Trees & Random Forests

Gaussian Mixture Models, EM Algorithm

(Hyper) Parameter Tuning

Simple Neural Nets

## Certificates

**Data Camp**: Data Scientist Track in R

**Data Camp**: Data Scientist Track in Python

**deeplearning.ai**

**fast.ai**

## Courses

Prob. & Stat. Inference I & II

Statistical Methods I & II

Econometrics

Applied Econometrics

Intro to Machine Learning

Intro to Data Science

Image Analysis

Cancer Genomics & Genetics

Statistical Genetics RNA-Seq