

# Brayton Hall

[braytonhall@gmail.com](mailto:braytonhall@gmail.com) | [Github](#) | [Data Science Blog](#) | [LinkedIn](#)

## Data Scientist | Data Engineer

Experience in data acquisition and data modeling, statistical analysis, machine learning, deep learning, and NLP. With a background in mathematical logic and philosophy of language at Virginia Tech, I excel at **communicating technical, academic, and scientific information in a way that is interpretable and precise**. My experience with data science, as well as my graduate experience with math, teaching, and writing is ideal for helping data driven companies make **informed, optimal decisions**.

## TECHNICAL SKILLS

Machine Learning, Deep Learning, NLP, Python, SQL, Docker, PySpark, Hadoop, scikit-learn, NumPy, Pandas, Keras, Regression, Classification, Clustering, Seaborn, matplotlib, Git, NLTK, Gensim, Word2Vec, BeautifulSoup, spaCy, Heroku, Streamlit, React, HTML, CSS, Javascript

## TECHNICAL PROJECTS

### Semantic Search Engine - [Github](#)

Created a 'search between the lines' app to search by connotations and misremembered quotes rather than exact fragments, inspired by the frustration of simple search engines on e-readers and websites

- Used Doc2Vec (a neural net package based on Word2Vec) in Gensim to vectorize 60,000 paragraphs from free ebooks on Project Gutenberg, and used them to train a model which can infer meaning based on cosine similarity
- Cleaned, organized, and visualized the data (1.2 million words) using Pandas and Seaborn, and stored the model and DataFrame on an S3 bucket with AWS
- Used a Docker container to deploy the app's front end on Heroku, which can be found on the project's Github page

### COVID-19 Twitter Topic Modeling - [Github](#)

Using a combination of natural language processing packages to find intelligible topic clusters on Twitter

- Scraped 100,000 'coronavirus' tweets using Twint and preprocessed them with Gensim, NLTK, and spaCy
- Used the Python package vaderSentiment to analyze tweets for non-neutral sentiments
- Created a Latent Dirichlet Allocation (LDA) model with Gensim to discover that five distinct Twitter communities emerged: 'Trump', 'Economy', 'Employment', 'Public Health', and 'UK/Boris Johnson'

### Tanzania Water Pump Access - [Github](#)

Predicted functionality of water pumps throughout villages in Tanzania in order to determine primary predictors of lack of access to clean water

- Geographically mapped non-functioning and functioning water pumps using Seaborn and Pandas
- Created a variety of models (Random Forests, KNN, Logistic, etc.) to predict broken pumps with 83% accuracy

## EMPLOYMENT HISTORY

Statistics Instructor, **The Learning Curve**, Charlotte, North Carolina 11/2017 - 08/2019

- Broke down complex statistics and math concepts for students at UNC Charlotte and in high school, identified individual strengths and weaknesses for improvement, and catered teaching style to students with different backgrounds
- Specialized in teaching linear algebra, probability theory, and statistics including normal distributions, z-scores, confidence intervals, t-tests, hypothesis testing, A/B testing

Graduate Teaching Assistant, **Virginia Tech**, Blacksburg, Virginia 08/2015 - 05/2017

- Managed meeting times and review sessions for exams, organized class logistics with the professor for about 60 students per semester, and simplified difficult concepts to create an engaging and approachable environment for intro students
- Created a grading database for professors, documented student papers and exams, and lectured a 150 person seminar

## EDUCATION

**Flatiron School**, New York, NY 01/2020 - 05/2020

Immersive Data Science Program

**Virginia Tech**, Blacksburg, Virginia 08/2015 - 05/2017

MA in Philosophy

- *Relevant coursework:* Advanced Logic, Metalogic, Philosophy of Science, Philosophy of Language

**UNC Chapel Hill**, Chapel Hill, North Carolina 08/2009 - 05/2013

BS in Psychology

- *Relevant coursework:* Statistics, Decision Theory, Calc I, Calc II, Symbolic Logic, Physics I, Cognitive Psychology