

Brayton Hall

Charlotte, NC 28202 | 704-450-6631 | braytonhall@gmail.com | [Github](#) | [Data Science Blog](#) | [LinkedIn](#)

Data Scientist | Data Engineer

Experience in data acquisition and modeling, statistical analysis, machine learning, and natural language processing. With a background in mathematical logic and philosophy at Virginia Tech and immersion in real large-scale ML problems, I excel at **building end-to-end ML pipelines that actually work** in production with big data. My focus has been on the fine-tuning of generative transformers for data labeling, classification, and summarization.

TECHNICAL SKILLS

Python, Docker, PyTorch, TensorFlow, Git, Spark, SQL, Hadoop, Impala, Hive, and LLMs such as BERT, GPT, T5, Pegasus

EMPLOYMENT HISTORY

Data Scientist, **Bank of America**, Charlotte, North Carolina 01/2022 - Present

- Develop and own the primary NLP models (T5, Pegasus, GPT) for abstractive text summarization between bank agents and customers, reducing agent manual note-taking by 3 minutes per call, resulting in **projected \$100 million cost reduction for complaint tracking**
- Train primary T5 model with 600,000 manually-written agent notes and 150 million speech-to-text converted conversations to produce highly precise summaries of BofA customer-agent interactions, which concern a vast range of products and complaint types for conversations lasting between 1 and 60 minutes
- Present interpretable findings to executives using transformer model output for cost reduction proof-of-concepts, showcasing **massive performance boosts on automated dissatisfaction and complaint capture** compared to outdated lexicon models, and document model development for Model Risk Management (MRM)

Data Scientist, **PrimeNeuro**, Durham, North Carolina 10/2020 - 04/2021

- Used ML algorithms on infant-brain fMRI imaging data to predict emergent behavior, with the aim of early interventions and diagnosis for high-risk infants
- Used techniques such as PCA, Shapley values, and tree-based models on massively wide datasets (100,000+ features) for feature extraction, feature reduction, and interpretable modeling for specific brain regions

Graduate Philosophy and Statistics Instructor, **Virginia Tech**, Blacksburg, Virginia 08/2015 - 08/2019

- Lectured 20 to 150 person seminars focused on logic, statistical inference, ethics, philosophy of natural science, and created and maintained a grading database for professors
- Taught statistics post-graduation, with a focus on applied concepts such as confidence intervals, t-tests, hypothesis testing, and A/B testing, linear algebra, probability theory, Bayesian statistics, and linear regression

TECHNICAL PROJECTS

Semantic Search Engine - [Github](#)

Created a 'search between the lines' app from scratch, 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 in Gensim to vectorize 60,000 paragraphs from free ebooks on Project Gutenberg and trained a model to infer meaning based on cosine similarity
- Cleaned, organized, and visualized the data using Pandas and Seaborn, stored on an S3 bucket with AWS, deployed using Docker on Heroku

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

Used a combination of NLP packages to find intelligible topic clusters on Twitter

- Scraped Twitter for high volumes of Covid-19 tweets using Twint and preprocessed them with Gensim, NLTK, and spaCy for an LDA model with Gensim to discover that five distinct Twitter communities emerged in early 2020

EDUCATION

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

Data Science Program

- *Relevant experience:* Python, SQL, Statistics, Machine Learning, Computer Science, Data Visualization

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

MA in Philosophy

- *Relevant coursework:* Metalogic and Number Theory, Philosophy of Science, Ethics, 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, Cognitive Psychology