

Miles C. Butler

Washington, DC | (734) 680-6525 | butlermiles83@gmail.com | <https://github.com/MilesButler2019> | www.linkedin.com/in/milesbutler

Professional Experience

Department of Justice

Washington, DC

Data Scientist

August 2023 - Present

- Led a comprehensive code analysis of a complex AI SaaS product (10+ million lines of Java + SQL code) using reverse engineering techniques. This analysis documented how the product works and identified potential anti-competitive behaviors.
- Provided expert insights for legal action against anti-competitive practices: Presented a clear and concise overview of the algorithm's functionality to diverse audiences, including lawyers, economists, and attorney generals
- Streamlined big data processing and enhanced document review: Successfully implemented Databricks into production, enabling efficient handling of large-scale data sets. Utilized cutting-edge Large Language Models (LLMs) to improve the search engine's capabilities, facilitating faster and more effective document review.

Meta (Formerly Facebook)

Menlo Park, CA

Data Scientist Intern

May 2022 - Aug 2022

- Designed and implemented an ensemble classifier using Logistic Regression, LightGBM, and XGBoost to optimize ad campaign performance for advertisers, achieving a remarkable 91% accuracy.
- Spearheaded the prototyping of an in-product explainability feature for XGBoost models using SHAP. This enhances user experience by providing advertisers with transparency into model predictions, leading to a 5% reduction in A/B testing time for ad creation.
- Built and Deployed a real-time dashboard leveraging Spark and SQL to analyze historical ad performance and compare effectiveness of ranking models. This enabled the identification of key performance metrics to optimize future ad campaigns.

AAA Insurance

Detroit, MI

Data Scientist Intern

May 2021 - Aug 2021

- Developed and implemented machine learning models using Logistic Regression, RNN, and Prophet to predict breakdowns in specific zip codes and times, increasing accuracy by up to 4% compared to existing models for Roadside Assistance; enabled improved demand forecasting and resource allocation for millions of tow trucks calls annually
- Built a full-stack application leveraging GraphSAGE to identify fraudulent members with 74% accuracy, contributing to enhanced risk management.

University of Michigan Alumni Donations

Ann Arbor, MI

Data Scientist Intern (NLP)

May 2020 - May 2021

- Developed a scalable Entity Resolution (ER) system using Spark GraphX on Databricks to consolidate information from 10+ databases for over 700,000 alumni, enabling effective targeting of donation campaigns.
- Improved existing ER accuracy by 16% by integrating techniques like graph and word embeddings, ultimately leading to a 25% increase in fundraising email clicks.
- Built a BERT classifier with 92% accuracy, categorizing unstructured data by topic which facilitated greater understanding of individual alumni interests.

Academic Projects

Nock Lab

Harvard University

- Modeled times series data with gaussian process regression to predict likelihood of a given person committing self harm through irregularly sampled survey data. Featured in New York Times article "Can Smartphones Help Predict Suicide?" on Sep 30, 2022.

Longitudinal analysis for Lung Disease Diagnosis: a CNN-RNN based approach

Massachusetts Institute of Technology

- Researched on a novel CNN-RNN framework to optimally capture both the spatial and longitudinal from the 3D images FMRI's with temporal features into a single classifier that provides 4% better results than the traditional CNN approaches.

Technical Skills

Programming Languages: Python, C++, C, Java, JavaScript, R, SQL, HTML, CSS

Database: MySQL, Postgres, CouchDB, Presto, Hadoop, Spark, Hive, MongoDB **Cloud:** Azure, AWS, GCP

Technologies/Frameworks: Pytorch, TensorFlow, Scikit-Learn, LLM, Maven, Git, Docker, REST API, GRCP, TCP/IP, HTTP

Other Skills: Tableau, Jira, Notion, Agile, Statistics, Microsoft Office, Teamwork, Written and Verbal Communication, Problem Solving

Education

Harvard University/ Massachusetts Institute of Technology (Dually Enrolled)

Cambridge, MA

Master of Science in Data Science

May 2023

University of Michigan

Ann Arbor, MI

Bachelor of Science in Information: Data Science and Analytics, James Scholar, University Honors

May 2021