Blake Law

blakelaw@gatech.edu
(404) 285-6750
github.com/blakelaw
blakelaw.dev

Education

Georgia Institute of Technology — Expected Dec 2024
Bachelor of Science in Mathematics, Data Science Concentration — GPA: 3.9

Experience

Perpay — June – Sep 2024
Incoming Data Science Intern — Philadelphia, PA

Georgia Institute of Technology — May 2023 – Present
Teaching Assistant — Atlanta, GA

- Provided feedback for exams and assignments in three advanced statistics and one PDE class

Kennesaw State University — Aug – Dec 2019
Research Intern — Kennesaw, GA

- Modeled traffic flow at three local intersections with 240-vehicle sample processed with pandas and Excel
- Discovered roundabouts had shortest wait time by 14 seconds through ANOVA and post-hoc tests in R
- Presented findings to panel of judges at research symposium, advocating for greater roundabout adoption due to higher safety profile (78-82% reduced fatalities) and efficiency

Projects

NBA Referee Neutrality Investigation – Python, pandas, Azure, SQL, R, SciPy, Tableau

- Created four methods to assess NBA referee bias, including propensity score matching and PCA approaches
- Cleaned and preprocessed 13M rows of game data in Azure SQL Database through CTEs and join operations
- Discovered zero systemic bias (all referees $p > 0.05$) for referees spanning 2004-2022 seasons

Identifying Microregions in Georgia – Python, (pandas, scikit-learn), OpenStreetMap

- Identified over 350 microregions (small unofficial regions) in Georgia through OpenStreetMap (OSM) data
- Preprocessed over 12 million data points from OSM representing local features using its REST API and pandas
- Created custom TF-IDF algorithm for similarity scores and used DBSCAN clustering for region identification

Duplicate Account Detection with NLP – Python, NLTK, pandas, scikit-learn, Matplotlib

- Developed three approaches for identifying duplicate accounts on discussion boards by writing style
- Preprocessed 400,000 forum comments over 13 months with PySpark, pandas, and Disqus Web API
- Achieved a true positive rate of 93% in identifying duplicate accounts with hierarchical cluster analysis

Quantifying Media Slant with LLMs – Python, PyTorch, pandas, AWS, Databricks

- Assessed media bias of three news sources through LLM outputs on Political Compass questions
- Fine-tuned LLaMA-7B model on 3000-article dataset using PyTorch on Databricks, backed by EC2 instances
- Found Fox News was the highest in economic polarization while NPR exhibited the least overall bias

Skills

Languages	Python, SQL (PostgreSQL, SQLite), R, MATLAB
Libraries	pandas, NumPy, scikit-learn, PyTorch, Matplotlib, NLTK, SciPy, TensorFlow
Tools	Tableau, Excel, Git, LaTeX, MongoDB, Linux, Docker, Kubernetes, dbt, MLFlow
ML / Stats	A/B Testing, Causal Inference, Regression, NLP, Gradient Boosting, Neural Networks
Big Data	AWS (S3, EC2, Redshift), Spark, Hadoop, Azure, GCP (Vertex), Databricks, Elasticsearch