

Josh Kraus

Data Scientist

joshtkraus@gmail.com
(919) 889-1180
Charlotte, NC
[LinkedIn](#) | [Portfolio](#)

Education

M.S. Analytics

Georgia Institute of Technology
2022 - 2024
GPA: 4.00, Summa Cum Laude

M.M. Education

University of Florida
2020 - 2021
GPA: 4.00, Summa Cum Laude

B.M. Performance

University of North Carolina -
Greensboro
2016 - 2019
GPA: 4.00, Summa Cum Laude

Skills

Programming

Git/GitHub

Python: NumPy, Pandas, PySpark,
scikit-learn

SQL: MySQL, Snowflake SQL

SAS

MLOps & Deployment

MLflow: Tracking, Registry

CI/CD: Unit Testing, Pipelines

Cloud Platforms: AWS, Snowflake,
Databricks

Machine Learning

Tree-Based Models: LightGBM, XGBoost

Deep Learning: Keras

Statistical Modeling: Regression, GLMs

Optimization: Optuna

Interpretability: SHAP, Feature
Importance

Visualization

Python: Matplotlib, Seaborn, Plotly

Dashboards: Power BI

Work Experience

Data Scientist

Liberty Mutual Insurance

Analyst I: May 2024 - Present

- Built custom AutoML package with automated hyperparameter tuning (Optuna), model interpretability (SHAP), and CI/CD deployment via GitHub Actions.
- Developed 10 GBMs for insurance pricing with a cross-functional team, achieving 3% Gini improvement using LightGBM and MLflow on Databricks.
- Generated \$20M in loss improvement by leading 3-person team to engineer 19 production features for actuarial pricing models, processing 50M+ rows using PySpark on AWS and Snowflake.
- Managed two cohorts of interns (4 total) across 12-week programs, leading technical onboarding, defining project scope, and overseeing execution and deliverables.
- Delivered \$1M in loss improvement by identifying and mitigating outlier losses across 10 actuarial models through automated data processing and modeling pipelines in PySpark and AWS.

Intern: June 2023 - May 2024

- Improved loss cost prediction by identifying and implementing new features in GLMs, strengthening statistical modeling accuracy for expense projections.

Founder & Analytics Consultant

Mindful Modeling

January 2022 - December 2023

- Optimized staffing allocation for a high-volume educational institution by analyzing demand patterns to improve resource utilization.
- Built custom Power BI dashboards incorporating time series forecasts and historical trends from 100,000+ student records to support staffing decisions for two student services organizations.

Music Educator

Green Level High School

August 2016 - November 2023

- Designed and delivered curriculum while collaborating with colleagues to establish learning objectives and measure student progress.
- Adapted instruction to meet diverse student needs, mentoring 50+ students and strengthening communication, planning, and feedback skills.

Conference Presentations

How Weather Events Lead to Overfitting in GBM & How to Fix it

Liberty Mutual Data Science Forum, Boston, October 2025

Notable Projects

College Basketball Bracket Prediction

Developed six ensemble models combining Gradient Boosting Machines (XGBoost) and Multilayer Perceptrons (Keras), using Optuna for hyperparameter optimization to predict NCAA Tournament outcomes. Scraped and curated 18 years of tournament data via BeautifulSoup, creating a dataset with 150+ predictive features. Achieved robust backtested performance over 12 years, correctly selecting the tournament winner 82% of the time.