Souvik Bag

Columbia, Missouri | sbk29@umsystem.edu | 929-276-8537 | github.com/souvik-bag | souvik-bag.github.io

Skills

Programming Languages: Python (Pandas, NumPy, Scikit-learn, XGBoost, Matplotlib, Seaborn), R, SQL

Deep Learning Frameworks: PyTorch, TensorFlow, Torchvision, OpenCV

Statistical Analysis: A/B Testing, Spatial Modeling, Hypothesis Testing, Bayesian Computation, Non-Linear Models

Experience

Applied Machine Learning Fellow — Los Alamos, NM
Los Alamos National Laboratory — May 2024 - August 2024

- Managed model risk by developing a **conformal prediction** framework in PyTorch to rigorously quantify the uncertainty of a Masked R-CNN model's output
- Boosted model reliability by **25%** over standard validation techniques and achieved a **96% prediction coverage rate**, ensuring high confidence for critical decision-making
- Owned the end-to-end model development lifecycle, from preprocessing and augmenting a specialized image dataset to fine-tuning different image segmentation models on High-Performance Computing (HPC) clusters

Graduate Research Assistant — Columbia, MO
University of Missouri — May 2023 - August 2023

- Engineered a massive time-series dataset for predictive modeling, processing over 10 years of raw data (3,650+ days) using R and SQL
- Developed a sophisticated deep learning forecasting model in PyTorch on HPC clusters to predict the occurrence, shape, and location of rare, high-impact events
- Innovated a custom loss function to solve a key challenge in the model's training, improving prediction accuracy for event shape and location by 30%
- Implemented uncertainty quantifying methods in Python to provide statistically rigorous confidence intervals for each forecast, enabling superior data-driven risk assessment

Education

Ph.D. in Statistics — August 2022 - May 2027
University of Missouri — GPA: 3.8/4.0

- **Coursework:** Spatial Statistics, Neural Networks, Computer Vision, Data Analysis, Statistical Computation & Simulation

Master of Science in Statistics — August 2019 - May 2021
Banaras Hindu University, India — GPA: 8.98/10.0

- **Coursework:** Statistical inference, Hypothesis testing, Stochastic processes, Measure theory, Advanced Bayesian analysis

Publications

- **Bag, S.**, Gupta, K., & Deb, S. (2024). *A review and recommendations on variable selection methods in regression models for binary data.* **Under review**
  Google Scholar

Projects

Binary Prediction of Smoker Status using Bio-Signals Github

- Engineered and selected features from 10+ bio-signals for over 5,000 subjects to build a robust classification model for predicting a key health outcome
- Built and benchmarked multiple classifiers, establishing a baseline with a logistic regression model in **R** and ultimately developing a champion **XGBoost** model with a 20% lift in performance (AUC: 0.881)
- Further explored advanced methods, including neural networks, to conduct a comprehensive evaluation and demonstrate breadth in modeling techniques

Analysis of Indian Super League Data Using Bivariate Poisson and Double Poisson Models Github

- Developed a **Bayesian hierarchical model** in R-Stan to quantify underlying performance drivers and isolate group-level effects within a complex system (e.g., team "home advantage")
- Executed **MCMC** simulations with over 6,000 iterations to generate robust posterior distributions for model parameters, demonstrating expertise in advanced statistical inference

Awards

NS-Ramaswamy Predoctoral Fellowship (Awarded to **8 of 1,000+ applicants**) July 2021 - June 2022
Indian Institute of Management Bangalore

Bert Winemiller Award for Poster Presentation (Data Science Research) October 2023
University of Missouri Columbia

Graduate Professional Council Travel Award (to present at **Joint Statistical Meetings 2024**) August 2024
University of Missouri Columbia

MUIDSI Hackathon, Placed in **Top 3 of 40+ Teams** (Project: AI Health Assistant using GPT) February 2025
University of Missouri Columbia

Silver Medal (**Ranked 2nd in graduating class**, Dept. of Statistics) August 2019
Midnapore College, India

Leadership

Secretary, *Statistics Graduate Student Association* August 2023 – Present
University of Missouri Columbia

- Spearheaded a new mentorship program, successfully pairing **15+ incoming graduate students** with senior peers to improve student onboarding and departmental retention
- Coordinated the planning and execution of **5+ professional development workshops** and networking events for a department of over 50 graduate students

Technical Mentor, *ASA DataFest Competition* April 2023, 2024
Hosted at University of Missouri Columbia

- Mentored **10+ student teams** from multiple universities during a 48-hour data analysis competition, providing expert guidance on statistical modeling, feature engineering, and data visualization
- Advised teams on translating ambiguous business problems into actionable analytical frameworks using a large, real-world dataset from a leading company

R Programming Peer Tutor, *Department of Statistics* August 2022 – May 2023
University of Missouri Columbia

- Served as the departmental technical expert for the R language, providing one-on-one consulting to **over 30 graduate students** on statistical coding for coursework and dissertation research
- Designed and led **3 workshops** on advanced R topics, including data manipulation with ‘dplyr‘ and data visualization with ‘ggplot2‘, enhancing the technical skills of the student community