Nasheed Jafri

Bloomington, IN | (260) 508-9096 | njafri@iu.edu | Webpage: nashjafri.github.io | LinkedIn: in/nasheedjafri

EDUCATION

Ph.D. Mathematics (Ph.D. Minor in Data Science) | *Indiana University, Bloomington, IN*	**GPA : 3.98**	Aug 2020 - May 2026
M.S. Applied Statistics | *Indiana University, Bloomington, IN*	**GPA : 4.00**	Jan 2024 - May 2026
M.S. Mathematics | *Indian Institute of Technology, Delhi, India*	**GPA : 3.80**	Aug 2018 - May 2020
B.S. Mathematics (Honors) | *University of Delhi, India*	**GPA : 3.82**	Aug 2014 - May 2017

PROFESSIONAL EXPERIENCE

INMAS - Internship Network in Mathematical Sciences, Urbana - Champaign, IL

Data Science Trainee — Sep 2024 - Feb 2025

- Analyzed Redfin housing data using multilinear regression in Python to model price variations across different U.S. cities and presented analytical findings to a technical audience.
- Performed EDA, hypothesis testing, and employed forward, backward, and stepwise model selection to predict wine quality based on physicochemical attributes from the Vinho Verde wine dataset.

Indiana University, Bloomington, IN

Associate Instructor in Linear Algebra for Data Science — Aug 2024 - May 2025

- Mentored students in foundational linear algebra concepts relevant to data science, including matrix operations, least squares, gradient descent, singular value decomposition, clustering and PCA.
- Designed interactive group learning sessions to reinforce practical applications in data analysis and ML.

Assistant Instructor in Probability and Statistics for Data Science — Jan 2024 - May 2024

- Developed curriculum to apply data analysis to real-world problems in social and natural sciences.
- Taught key concepts including statistical inference, hypothesis testing, maximum likelihood estimation, central limit theorem, bootstrap resampling, chi-square tests, and their applications in data analysis.

REU Mentor — Aug 2022 - Dec 2022

- Supervised a group of undergrad students in a graduate-level research project on Fourier Transform.

PhD Research - Linear Algebra and Matrix Theory — Jan 2021 - May 2026

- Developed a novel algorithm to construct invariant subspaces of nilpotent matrices using tableaux.
- Applied combinatorial methods to Linear Algebra and Matrix Theory, proving uniqueness of the algorithm- constructed invariant subspaces using discrete structures called puzzles and honeycombs.

PROJECTS

Loan Default Prediction for Home Credit — Aug 2024 - Dec 2024

- Collaborated with a team of data scientists to predict loan defaults for Home Credit using Machine Learning in Python on large datasets containing 300k to 27 million samples, achieving 92% test accuracy.
- Led the EDA and feature engineering phase, handled missing values and performed correlation analysis.
- Built and optimized models (Logistic Regression, Decision Trees, Random Forests, Gradient Boosting, XGBoost, SVC and MLP neural networks) using PCA and ensemble methods like voting and stacking.

Link
GitHub Repository

Approximate Bayesian Computation for Disease Outbreak — Aug 2024 - Dec 2024

- Implemented Approximate Bayesian Computation in R to fit an epidemic model for influenza outbreaks.
- Built custom functions for parameter sampling, data simulation, and ABC rejection sampler algorithm.
- Performed model comparisons by estimating posterior probabilities to analyze variations in infection transmission rates across outbreaks of the same strain and different strains of the virus.

Link
GitHub Repository

SKILLS

- **Programming:** Python, R, SQL, C
- **Database Technologies:** MySQL, PostgreSQL, pgAdmin, MySQLWrokbench
- **Web Technologies:** HTML, CSS, MAMP
- **Tools/Platforms:** Jupyter, Google Colab, GitHub (Version Control), R Studio, Conda, Docker
- **Python Libraries:** Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, Statsmodels, Matplotlib, Seaborn, SciPy, LightGBM, XGBoost
- **Data Analysis:** EDA, Data Cleaning/Preprocessing, Feature Engineering, Feature Selection, Dimensionality Reduction, Data Aggregation
- **Machine Learning:** Regression (Linear, Logistic, Ridge, LASSO), Classification (SVM, Decision Trees, Naive Bayes, k-NN), MLP, Neural Networks, Random Forest, Clustering, Principal Component Analysis, Hyperparameter Tuning, Cross-Validation, ML Pipelines
- **Statistics:** Generalized Linear Models, Hypothesis Testing, Bayesian Inference, ANOVA, Model Selection, Monte Carlo Simulations
- **Mathematics:** Linear Algebra, PDEs, Numerical Analysis, Graph Theory, Dynamical Systems, Linear Programming, Optimization

COURSEWORK

Machine Learning (A+), Econometrics (A+), Statistics (A+), Database Systems (A+), Applied Linear Models (A), Statistical Computing (A)