

# Nasheed Jafri

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

## EDUCATION

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

## PROFESSIONAL EXPERIENCE

<i>INMAS - Internship Network in Mathematical Sciences, Urbana - Champaign, IL</i>		
<b>Data Science Trainee</b>		Sep 2024 - Feb 2025
<ul style="list-style-type: none"><li>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.</li><li>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.</li></ul>		
<i>Indiana University, Bloomington, IN</i>		
<b>Associate Instructor in Linear Algebra for Data Science</b>		Aug 2024 - May 2025
<ul style="list-style-type: none"><li>Mentored students in foundational linear algebra concepts relevant to data science, including matrix operations, least squares, gradient descent, singular value decomposition, clustering and PCA.</li><li>Designed interactive group learning sessions to reinforce practical applications in data analysis and ML.</li></ul>		
<b>Assistant Instructor in Probability and Statistics for Data Science</b>		Jan 2024 - May 2024
<ul style="list-style-type: none"><li>Developed curriculum to apply data analysis to real-world problems in social and natural sciences.</li><li>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.</li></ul>		
<b>REU Mentor</b>		Aug 2022 - Dec 2022
<ul style="list-style-type: none"><li>Supervised a group of undergrad students in a graduate-level research project on Fourier Transform.</li></ul>		

## SKILLS

<ul style="list-style-type: none"><li><b>Programming:</b> Python, R, SQL, C</li><li><b>Web Technologies:</b> HTML, CSS, MAMP</li><li><b>Python Libraries:</b> Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, Statsmodels, Matplotlib, Seaborn, SciPy, LightGBM, XGBoost</li><li><b>Data Analysis:</b> EDA, Data Cleaning and Preprocessing, Feature Engineering, Feature Selection, Dimensionality Reduction, Data Aggregation</li><li><b>Machine Learning:</b> Regression (Linear, Logistic, Ridge, LASSO), Classification (SVM, Decision Trees, Naive Bayes, Discriminant Analysis, k-NN), Neural Networks, Random Forest, Clustering, Principal Component Analysis, Hyperparameter Tuning, Cross-Validation, ML Pipelines</li><li><b>Statistics:</b> Generalized Linear Models, Hypothesis Testing, Bayesian Inference, ANOVA, Model Selection, Monte Carlo Simulations</li><li><b>Mathematics:</b> Linear Algebra, PDEs, Numerical Analysis, Graph Theory, Dynamical Systems, Linear Programming, Optimization Methods</li></ul>	<ul style="list-style-type: none"><li><b>Database Technologies:</b> MySQL, PostgreSQL, pgAdmin, MySQLWrokbench, Snowflake</li><li><b>Tools/Platforms:</b> Jupyter, Google Colab, GitHub (Version Control), R Studio, Conda, Docker</li></ul>
---	--

## RELEVANT COURSEWORK

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

## PROJECTS AND RESEARCH

<b>Loan Default Prediction for Home Credit using Machine Learning</b>	Aug 2024 - Dec 2024
<ul style="list-style-type: none"><li>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.</li><li>Led the EDA and feature engineering phase, handled missing values and performed correlation analysis.</li><li>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.</li></ul>	
<b>Approximate Bayesian Computation for Disease Outbreak</b>	Aug 2024 - Dec 2024
<ul style="list-style-type: none"><li>Implemented Approximate Bayesian Computation in R to fit an epidemic model for influenza outbreaks.</li><li>Created custom functions for parameter sampling, data simulation, and ABC rejection sampler algorithm.</li><li>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.</li></ul>	
<b>PhD Research - Combinatorial Methods in Linear Algebra and Matrix Theory</b>	Jan 2021 - May 2026
<ul style="list-style-type: none"><li>Developed a novel algorithm to construct invariant subspaces of nilpotent matrices using tableaux.</li><li>Applied combinatorial methods to Linear Algebra and Matrix Theory, proving uniqueness of the algorithm-constructed invariant subspaces using discrete structures called puzzles and honeycombs.</li></ul>	

