# SHAKIL RAFI Ph.D.

Applied Mathematician
Data Scientist & M.L. Engineer
🗺 Fayetteville, AR, 72701, willing to relocate

sarafi@uark.edu ✉
+1 479 301 1866 📞
https://LinkedIn.com/in/shakilrafi 🔗
https://GitHub.com/2shakilrafi 🔗
https://2shakilrafi.github.io 🌐

Passionate and hard-working individual working at the intersection of machine learning, deep learning, and differential equations. I aim to use the tools of mathematics to bring value and make a difference in my community.

**Experience**:

(i) *Post-doctoral Fellow,*
*University of Arkansas Department of Agriculture*
*Little Rock, AR.*
*Aug. 2024 - Present*

- Wrangle terabyte scale genomic datasets, including denoising and demultiplexing.
- Use advanced machine learning tools like decision tree regression, heirarchical clustering, and deep learning to gain predictive insights from genomic data
- Collaborate and serve as external consultant to various parties across academia and industry. Some of the clientsinclude: Cobb Genetics, Inc., UAB Heersink School of Medicine, and USDA Food Safety testing team.

(ii) *Lecturer,*
*Department of Data Science,*
*Sam M. Walton College of Business,*
*University of Arkansas*
*Fayetteville, AR*
*Jan. 2023-May 2024*

- I T.A. a course on optimization methods focusing on `CVXPY` and `Pyomo`, as well as oversee optimization projects as part of the finals for the course.
- I also teach a course on the Fall on Introductory Programming Languages for Data Science focusing on introducing freshmen to the bash shell, Python and R with an emphasis on `Numpy`, `Pandas`, `tidyverse` including `dplyr` and `ggplot2`. Students were expected to submit homework using `Git`.

(iii) *Intern,*
*Health Economics,*
*Arkansas Blue Cross and Blue Shield*
*Springdale, AR,*
*May 2023–August, 2023*

- Worked with the health economics team to develop a Markov model of childbirth and pregnancy. We looked at over $40,000$ of our female customers and created a model to track the stages of pregnancy, from conception of child to bad pregnancy outcomes to bad birth outcomes to bad post-birth outcomes.

- Developed gradient boosted decision tree and neural network models using `Scikit-learn` to predict successful/unsuccessful child birth outcomes based on the pre-existing health conditions and substance abuse status of our users with $> 90\%$ accuracy.

- Developed random forest and neural network models using `Scikit-learn` to predict successful/unsuccessful child birth outcomes based on the demographics (median income, educational attainment, etc.) of our users with $> 90\%$ accuracy.

- Trained an LSTM, RNN network using `Keras` and `Tensorflow` to predict claims volumes based on claims data going back to 2018.

- Required to share work every Wednesday with the team using Shiny dashboards, and code was shared via Posit Workbench.

(iv) *Senior Graduate Assistant*
*Department of Mathematical Sciences*
*University of Arkansas*
*Fayetteville, AR.*
*Jan 2020 - Dec 2022*

- Taught pre-calculus, Calculus, Linear Algebra, Differential Equations and intro Statistics.
- Planned lessons and liaison-ed with the course coordinator in a team of five other instructors.

(v) *Graduate Teaching Assistant,*
*Department of Mathematics,*
*University of Arkansas.*
*Fayetteville, AR.*
*Aug. 2016 – Dec. 2019.*

- Graded and held office hours for classes of 30+ students
- Designed courses and planned lessons on advanced linear algebra linear algebra.

(vi) *Asst. Business Analyst*
*R.S.S. Wears Ltd.*
*Gazipur, Bangladesh.*
*Jun. 2015 - Jul. 2016.*

Built out and maintained the technology stack, including inventory, payroll, website, order tracking, and bookkeeping for a medium garment manufacturing start-up of about 150 employees.

(vii) *Concessions Manager,*
*Sodexo, Inc.*
*Troy, AL.*
*Aug. 2011 - May 2013.* Set up concessions for Troy University games, with a team of six other people, including logistics, safety, and customer satisfaction.

## Education

(i) *Ph.D. Applied Mathematics, 2020-2024*
*University of Arkansas, Fayetteville, AR*, cGPA: 3.65

- Recipient: SIAM Student Travel Awards
- Recipient: Lawrence Jesser Toll Jr. Endowed Scholarship
- Treasurer, Graduate Student Colloquium

Wrote a dissertation on the use of neural networks (feed forward multi-layer perceptrons) to solve a highly modified version of the heat equation. Along the way greatly simplified and modified a stochastic Monte Carlo process, give parameter estimates for more than a handful of novel neural networks and use these novel neural network architectures to show that they can be used to approximate the differential equation with parameter bounds that are a polynomial on the accuracy $\varepsilon$.

Parameter bounds for neural networks are of crucial importance to the literature. Smaller neural networks that can still approximate complex functions means that the technology of neural networks becomes more accessible in smaller computing devices. This also contributes greatly to minimizing energy consumption required to train these models which is of utmost importance in an AI-driven but increasingly climate challenged future.

(ii) *M.S. Pure Mathematics, 2016-2019*
*University of Arkansas, Fayetteville, AR*, cGPA: 3.55

- Recipient: Bangladesh-Sweden Travel Fund
- Vice-president: Bangladesh Student Organization, University of Arkansas

(iii) *B.S. Pure Mathematics, Minor: Philosophy*
*Troy University, Troy, AL*, cGPA: 3.55

- Recipient: Millenium Scholar's Award, a full-ride (tuition, room, and board) scholarship to Troy University.
- Member: The Alabama Eta Chapter of the national math honor society, Pi Mu Epsilon.
- Dean's List: Multiple semesters.

## Papers and Working Papers

(i) Rafi S., Padgett J. L., Nakarmi U. Towards an Algebraic Framework For Approximating Functions Using Neural Network Polynomials. arXiv. `https://doi.org/10.48550/arXiv.2402.01058`

We make the case for neural network objects and extend an already existing neural network calculus. Our aim will be to show that, yes, indeed, it makes sense to talk about neural network polynomials, neural network exponentials, sine, and cosines in the sense that they do indeed approximate their real number counterparts subject to limitations on certain of their parameters, q, and $\varepsilon$. While doing this, we show that the parameter and depth growth are only polynomial on their desired accuracy (defined as a 1-norm difference over $\mathbb{R}$), thereby showing that this approach to approximating, where a neural network in some sense has the structural properties of the function it is approximating is not entirely intractable.

Currently under review for at *The Journal of Artificial Intelligence Research.*

(ii) Cory Gillette, Halley Fritze & Shakil Rafi. *Categories Are All You Need (Forthcoming)*.

Recent work by Hutzenthaler, Jentzen, and Salimova describe neural networks as an ordered tuple of ordered pairs. Along with the realization operator, sums, and scalar multiplication this is sufficient to describe even architectures as sophisticated as Transformers. We conjecture that the set of neural networks described forms a weak bi-module with realization being a functor from this weakly described category to the category of sets with homeomorphisms.

(iii) Rafi, S. Gender Disparities in Arkansas, and Income Disparities in the US for the PPP Loan Program. Preprints 2023, 2023090654. `https://doi.org/10.20944/preprints202309.0654.v1`

We look at data from the Paycheck Protection Program, a pandemic-era measure to help small businesses meet essentials like payroll. We showed that within zip-codes which are assumed to hav uniform median income, business size, race, and gender are are non-trivial predictors of the size of initial loan approval amount.

(iv) Rafi, Shakil; Nithila, Arna Nishita (2022). Who rides Uber anyway? A census-tract level analysis and clustering of ride-shares for the city of Chicago during the era of the pandemic. TechRxiv. Preprint. `https://doi.org/10.36227/techrxiv.21076042.v2`

We looked at data from the City of Chicago transportation providers datasset for the entire year of 2020. We followed methodology from Soria, Chen, and Stathoupolos, 2019 and sought to use K-means to cluster census tracts and glean information about customer behavior. We say that richer census tracts used ride-shares (Uber and Lyft) much more differently than poorer tracts and that median age of a census tract had little to do with Uber usage.

Submitted to *The European Journal of Transportation and Infrastructure Research*.

(v) Rafi, S. An Proposed Algorithm for Generating Criteria Necessary to Establish Congruence between Two Convex N-Sided Polygons in Euclidean Geometry. Preprints 2023, 2023030490. `https://doi.org/10.20944/preprints202303.0490.v1`

Undergrad research work focusing on geometrical congruence between n-gons. Presented at the MAA MathFest 2014 in Portland, OR.

## Presentations

(i) Rafi S. *Maximal Parameter Estimates for Neural Networks and Uncertainties in Approximation*. SIAM Conference on Uncertainty Quantification '24. Trieste, Italy.

(ii) Rafi S. *SOM-where in Chicago: Self Organizing Maps as a Way to Analyze Transportation Data*. OAK Fall 2023 Conference, Little Rock, AR.

(iii) Rafi S. *Machine Learning Through the Centuries*. SIAM Student Chapter Meeting. Fayetteville, AR. September 2023.

(iv) Rafi S. *A Clustering Analysis of Ride-shares for the City of Chicago during 2020*. SIAM Mathematics of Data Science Conference, San Diego, 2022

(v) Rafi S. *Decision tree based regression and the PPP loans. Were women and minorities worse off?.* Arkansas Summer Research Institute, 2022.

(vi) Rafi S.*Who rides Uber anyway? A census-tract level analysis and clustering of ride-shares for the city of Chicago during the era of the pandemic.* Arkansas Spring Lecture Series, 2022.

(vii) Rafi S.*Turing Machines and the Halting Problem.* MAA Regional MathFest, Troy University, 2015

(viii) Rafi S.*Polygonal Congruences: A Suggested Algorithm.* MAA Annual Mathfest. Portland, OR. 2014.

## Technical Skills

- `MATLAB`. Highly skilled in `MATLAB` especially in its use in classical numerical analysis algorithms (Cholesky, Gram-Schmidt, symmetric Lanczos, etc.)

- `Python`. Highly skilled in the use of Python especially libraries such as `Pandas, NumPy, Scikit-learn, Keras/TensorFlow, pyTorch, matplotlib, Plotly, Folium, leaflet.js, and GeoPandas`.

- `R/tidyverse`. Highly skilled in the use of `tidyverse` especially `dplyr` and `ggplot2` in the case of data science, and Posit Workbench and `Shiny` for data exposition.

  Also the author of the `R` package `nnR`. Currently serving as reviewer for the Journal of Open Source Software.

- `Mathematica`. Proficient in Mathematica. Certified Wolfram Technology Associate. Authored a course on Udemy on the use of Mathematica 9 for college level mathematics.

- `C/C++`. Proficient in `C` in the context of implementing various classical algorithms (Newton-Raphson, Euler method, Runge-Kutta etc.). Passing familiarity of `C++`.

- `Julia`. Proficient in the use of `Julia` as it is used in simulations for my dissertation.

- `SQL`. Familiarity with SQL. Certified in SQL as part of the IBM Data Science Certification.

## Workshops

- **MSRI Summer Graduate School in Machine Learning:** In the summer of 2023, I was also nominated by my department head Dr. Andrew Raich to attend the Mathematical Sciences Research Institute (MSRI, now known as Simons Laufer Mathematical Sciences Institute), Summer Graduate School on Machine Learning at UC San Diego. I mentored under Dr. Mikhail Belkin of the Halicioglu Data Science Institute focusing on over-parametrization of artificial intelligence models and why that might be a good thing, and Dr. Yusu Weng on adversarial patterns in image recognition.

- **ASRI Summer School in Data Science:** In the summer of 2022 I attended the Arkansas Summer Research Institute hosted at the Arkansas School of Mathematics, Sciences, and Arts. This is an NSF EPSCOR sponsored workshop that brought students from across nine different states to develop skills needed for Data Science. I looked at the paycheck protection program (PPP) during the pandemic and using gradient boosted decision trees showed that

black and women-owned businesses tended to be approved for smaller initial loan amounts even accounting for business size and median income of the zip code. Mentored under Dr. Karl Schubert of the Arkansas Data Science Advisory Council.

## Certifications

- *Generative AI: Working with Large Language Models*. Issued by LinkedIn Learning. 2024.

- *Fundamentals of Deep Learning*. Issued by the NVidia Deep Learning Institute. 2023.

- *IBM Data Science Professional*. Issued by IBM, Inc. in collaboration with Coursera, Inc. 2021.

- *Advanced Google Analytics*. Issued by Google, Inc.

- *Certified Wolfram Technology Associate*. Issued by Wolfram Research, Inc. 2014.

## Outreach Activities

- Participant and booth presenter, Arkansas Math Discovery Day 2024.

- Lead Judge, Computer Science Category, Northwest Arkansas Regional Science and Engineering Fair. 2023, 2024.

- Secretary, Graduate Student Colloquium. Department of Mathematical Sciences. 2022–Present.

- Member, Democratic Socialists of America, Northwest Arkansas Chapter, 2020-2024

- Vice President, Bangladesh Student Organization, University of Arkansas, 2019-2020.

- Presenter, Department of Mathematics Celebration of Mind Festival. 2018, 2019.

- Foundation Member, Libertarian National Committee. 2015-2016.