

Saipranav Avula

sai.avula13@gmail.com | 908-798-8587 | [linkedin.com/in/sai-avula](https://www.linkedin.com/in/sai-avula) | Top Secret Security Clearance

Education

University of Michigan - Ann Arbor, MI

Sep 2024 - Dec 2025

Master of Science in Data Science - GPA: 3.85/4.0

- **Graduate Student Instructor** - Computing for Justice
- **Coursework** - Statistical Machine Learning, Probability Theory, Applied Machine Learning, Regression Analysis

Rensselaer Polytechnic Institute - Troy, NY

Sep 2019 - Dec 2022

Bachelor of Science in Computer Science - GPA: 3.76/4.0

- **Coursework** - Data Structures, Algorithms, Information Retrieval, Graph Theory

Technologies

Languages: Python, R, SQL, C++, Java

Frameworks/Tools: React, Git, Linux, AWS, Airflow, MLFlow, Docker, Postgres

Modeling: Classification, Regression, Time-Series, Deep Learning, Neural Networks, Neural Processes

Libraries: Scikit-learn, PyTorch, OpenCV, NumPy, Matplotlib, dplyr, ggplot2

Experience

Cooperative Institute for Great Lakes Research

Jan 2025 - Present

Data Science Researcher

Ann Arbor, MI

- Develop open-source package to train convolutional neural processes on multi-modal environmental Great Lakes data for downscaling applications and surface temperature uncertainty quantification.
- Utilize PyTorch and HPC to predict ideal buoy sensor placements to enhance weather forecasting.

MITRE Corporation

Apr 2023 - Aug 2024

Data Scientist - Computational Social Science

McLean, VA

- Applied DNN and XGBoost models to predict veteran benefits case duration with 87% AUC. Built end-to-end experimentation pipeline to mitigate data leakage and evaluate model performance on suite of metrics.
- Built ETL infrastructure for statistical outlier detection to identify under-performing federal training centers. Created data visualizations to support funding recommendations in report to Congress.
- Integrated PySpark data pipeline and React app to scale global vessel tracking system analysis window by 85%. Tool used by national security agencies to analyze global risks and respond to crises in real-time.
- Developed novel R NLP algorithms for US Census to identify data re-identification attacks. Collaborated with differential privacy domain experts and integrated into cloud hosted pilot service.

Janssen Pharmaceuticals of Johnson & Johnson

Jan 2022 - July 2022

Digital Innovation Data Scientist

New Brunswick, NJ

- Implemented data stream of research proposals to identify breakthrough technologies for bio-tech startup incubator to reduce manual data entry by 50%.
- Synthesized complex SQL queries and AWS integrations for data warehouse to enhance investment portfolio analysis.

Rensselaer Institute for Data Exploration and Applications

May 2020 - Aug 2020

Undergraduate Researcher

Troy, NY

- Leveraged graph mining algorithms to design COVID-19 risk dashboard to quantify social distancing policy efficacy in potentially reducing campus infection spread.
- Simulated 4000 realistic schedules from anonymized student data to generate contact networks and compare plans of distributing students among testing centers.
- Collaborated with health officials to dynamically update models based on new COVID-19 guidance. Models were used to inform successful policies which kept campus infection rates below regional average.