

ARYAN MEENA

Boston, MA | (617) 861-1307 | araj7042@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

SUMMARY

Data scientist who takes models end to end and cares most about whether a result is real. Set a fraud classifier threshold on the cost of a false decline at 0.96 precision, then killed a best-performing research result when it failed multiple-testing correction and published that instead.

EXPERIENCE

Boston University — Boston, MA

Sep 2025 – May 2026

Graduate Research Assistant, Quantitative Research

- Sourced and validated 3,932 trading days across 7 correlated financial series (2010 to 2025) into reusable Python modules, then evaluated a pre-specified 504-configuration grid of 24 rules, 7 instruments and 3 cost regimes
- Built the statsmodels validation layer (Newey-West HAC, Deflated and Probabilistic Sharpe Ratios, 5,000-resample block bootstrap, permutation tests, White's Reality Check) and reported the best result failing correction at Deflated Sharpe 0.046
- Isolated the one component that did clear correction at 0.405, and showed that ranking candidates by gross performance systematically selects those least able to survive real execution costs

Unified Mentor — Remote

Oct 2024 – Jan 2025

Data Analyst Intern

- Cleaned and joined 4.5M+ transactional records across inconsistent schemas, surfacing 2 distorted headline metrics
- Built a fraud classifier in scikit-learn (random forest, SMOTE, class weighting) at 0.96 precision and 0.80 recall on 284K transactions, setting the threshold on the cost of a false decline
- Delivered an ensemble demand forecaster at 8.4% MAPE with time-series cross-validation, monitoring for drift
- Automated the recurring reporting cycle (40% less manual effort) across Tableau and Power BI views over \$577K

Space Applications Centre (SAC), ISRO — Ahmedabad, India

Jun 2022 – Jul 2022

Machine Learning Research Intern

- Trained U-Net and DeepLabV3+ road-segmentation models in PyTorch (ResNet-50, Dice loss, IoU) on 6,723 aerial images, with QGIS in the geospatial loop
- Found that benchmark-trained models did not transfer to Indian satellite imagery, and traced why

PROJECTS

WikiFlow: Distributed ML on 60M Events | PySpark · GCP Dataproc · Keras

[GitHub](#) · [Paper](#)

- Engineered 13 leakage-free behavioral features across 60M revision events (1.16M editors, 25 GB) on Dataproc
- Wrote logistic regression and K-Means from scratch on PySpark RDDs; Elephas Keras net at 0.909 AUC vs 0.845

MedAI: 3D Medical Imaging Regression | PyTorch · MedicalNet · FiLM

[GitHub](#)

- Fine-tuned MedicalNet ResNet-34 with FiLM conditioning on multi-tracer 3D volumes to MAE 12.67 (r 0.968)
- Placed 3rd of 40+ teams as technical lead of 6, freezing preprocessing and evaluation interfaces first

Hospital Readmission Risk and Cost Model | Python · scikit-learn · SHAP

[GitHub](#)

- Benchmarked 10 algorithms on 30K records at a 7:1 imbalance, tuning the threshold on cost and adding SHAP
- Projected \$1.77M annual net savings per hospital, with break-even at 21.3% intervention effectiveness

SKILLS

Languages and Data: Python, SQL, R, C++, PostgreSQL, pandas, NumPy, SciPy, statsmodels, PySpark

Machine Learning: scikit-learn, XGBoost, PyTorch, Keras, random forest, gradient boosting, SHAP, SMOTE

Statistics: hypothesis testing, A/B testing, ANOVA, regression, time series, cross-validation, bootstrap

Cloud and Tools: GCP (Dataproc, BigQuery), Docker, Git, Linux, LaTeX, Power BI, Tableau, Matplotlib

PUBLICATIONS

- Meena, A., et al. *WikiFlow: Predicting Wikipedia Editor Dropout at Scale*. EAI CSECS 2026, Springer LNICST
- *Accelerated CNN Training with Genetic Algorithm Optimization*. IEEE IATMSI 2024, IEEE Xplore
- *Enhanced Scene Text Recognition for Geocoding and Accessibility*. ICADIE 2024, Springer
- **First author on all three**; 1 further manuscript under review

EDUCATION

Boston University, Metropolitan College — M.S., Applied Data Analytics | GPA: 3.75/4.0

Jan 2025 – May 2026

Indian Institute of Space Science and Technology — B.Tech., Electronics and Communication

Aug 2019 – May 2023