

# Aryan Meena

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

## SUMMARY

Data Scientist with an M.S. in Applied Data Analytics who takes models end to end, from stakeholder requirements and messy multi-source data through feature engineering, validation and deployment. I work in Python, SQL, scikit-learn and PySpark across healthcare, finance and web-scale data, and I care most about turning a model's numbers into a decision someone can act on. Author of two published peer-reviewed papers, with a third accepted at EAI CSECS 2026.

## EXPERIENCE

### Boston University — Research Assistant, Quantitative FX Research

Boston, MA • Sep 2025 – May 2026

Stack: Python • pandas • NumPy • statsmodels • Jupyter

- Sourced, cleaned and validated 15 years of daily market data across the U.S. Dollar Index and six correlated forex pairs, building the reusable Python modules the analysis ran on.
- Designed a 24-rotation intraday return-decomposition framework from scratch; the top rotation returned 219% against a 22% buy-and-hold benchmark at a 1.06 Sharpe in backtesting, with a derived momentum signal holding across all six pairs.
- Stress-tested every result across transaction-cost regimes using ANOVA and hypothesis testing, since a return that large usually means the test is wrong, then co-authored a manuscript now under review at *Traders Magazine*.

### Unified Mentor — Data Analyst Intern

Remote • Oct 2024 – Jan 2025

Stack: Python • scikit-learn • SQL • Tableau

- Gathered modeling and reporting requirements from product and operations stakeholders, then cleaned and joined 4.5M+ transactional records, resolving nulls, duplicates and schema inconsistencies.
- Built a credit-card fraud classifier (Random Forest with SMOTE and class weighting) reaching 0.96 precision and 0.80 recall across 284K transactions, tuning the precision-recall tradeoff around the cost of a false decline.
- Built an ensemble demand forecaster at 8.4% MAPE using time-series cross-validation, owned the recurring reporting cadence and automated it to cut manual effort by an estimated 40%, delivering self-service Tableau views across \$577K in revenue and 4.9K orders.

### Indian Space Research Organisation (ISRO) — Machine Learning Research Intern

Ahmedabad, India • Jun 2022 – Aug 2022

Stack: Python • PyTorch • transfer learning • Docker

- Trained a VGG-16 transfer-learning CNN on multi-spectral satellite imagery (82.03% accuracy on raw, non-curated images), then packaged it into a Dockerized Linux inference pipeline that cut manual annotation effort by an estimated 60% and was adopted by the research team.

## PROJECTS

### Hospital Readmission Prediction | scikit-learn • SMOTE • SHAP

[GitHub](#)

- Worked with 30K discharge records at a 7:1 imbalance, using SMOTE and 5-fold stratified CV across ten algorithms. When the signal came out weak, I reframed the problem around what a hospital would do: tuned the threshold for recall, used SHAP so clinicians could see the reasoning and built a cost-benefit model projecting \$1.77M in annual savings.

### DXY FX Strategies | Python • pandas • backtesting • ANOVA

[GitHub](#)

- Backtested 24 intraday strategies on the U.S. Dollar Index and six FX pairs over 15 years; the best turned \$100 into \$318.80 against \$121.72 buy-and-hold at a 1.06 Sharpe versus 0.22, tested at 0, 1 and 2 bps transaction costs to confirm the edge survived realistic friction.

### Jet Engine Remaining Useful Life | Python • R • time-series

[GitHub](#)

- Built predictive-maintenance models on NASA C-MAPSS turbofan sensor data, framing engine failure as both a classification and a remaining-useful-life regression problem, reaching AUC 0.945 on failure prediction.

## PUBLICATIONS & AWARDS

- **Meena, A., Patel, K., Alizadeh-Shabdiz, F.** *WikiFlow: Predicting Wikipedia Editor Dropout at Scale Using Distributed Machine Learning on Google Cloud*. EAI CSECS 2026, Springer LNICST (accepted). 60M events, 1.16M editors on GCP Dataproc; AUC 0.909. [GitHub](#) • [Paper](#)
- **MedAI Hackathon — 3rd of 40+ teams** (2026): Amyloid PET Centiloid prediction; MedicalNet ResNet-34 with FiLM conditioning in PyTorch, test MAE 12.67 CL (r = 0.968). [GitHub](#)
- **Meena, A., et al.** *Accelerated CNN Training with Genetic Algorithm Optimization*, IEEE IATMSI 2024 • *Enhanced Scene Text Recognition for Accessibility*, ICADIE (Springer) 2024.

## EDUCATION

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

Boston, MA • Jan 2025 – May 2026

Coursework: Advanced Machine Learning, Foundations of ML, Data Mining, Big Data Analytics, Data Science with Python, Web Mining & Graph Analytics

Teaching Assistant: MET CS 767 (Advanced ML) & MET CS 577 (Data Science with Python), Jan–May 2026; supported 60+ graduate students.

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

Kerala, India • Aug 2019 – May 2023

Coursework: Machine Learning for Signal Processing, Probability & Statistics, Computer Vision

## TECHNICAL SKILLS

**Languages & ML:** Python, SQL, R, scikit-learn, XGBoost, PyTorch, Random Forest, feature engineering, SHAP, SMOTE

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

**Data & Cloud:** PySpark, PostgreSQL, GCP (BigQuery, Dataproc), AWS (S3, Lambda), Docker, Git