

AADITYA VARDHAN DAS

+91 9899336502 ◇ Gurugram, India

avardhan.das@gmail.com ◇ aadityavardhan.com

SUMMARY

Data Scientist with 1.5+ years of experience building production-grade machine learning solutions across rare-disease patient identification, HCP targeting, patient event prediction, and healthcare analytics. Experienced in leading end-to-end data science engagements, developing ML systems over multi-million-patient longitudinal datasets, and productionizing scalable pipelines on Azure Databricks. Currently leading a cross-functional team of 2 Data Scientists and 2 Business Analysts while serving as the primary data science point of contact for client stakeholders.

EXPERIENCE

Data Scientist - 1

ProcDNA Analytics

Jan 2026 – Present

Gurugram, India

LGMDR9 Patient Identification & Launch Analytics

- Led Phase 1 of an end-to-end ultra-rare disease patient identification initiative supporting launch readiness for the only therapy in the indication; managed a cross-functional team of **2 Data Scientists and 2 Business Analysts** and served as the primary DS point of contact for senior client stakeholders.
- Designed the cohort strategy across **204 confirmed/high-confidence R9 patients, 3,030 patients with alternate LGMD subtypes**, and **19,166 unlabeled LGMD patients**; constructed a higher-confidence subset of **1,779 reliable negatives** to minimize subtype misclassification during model development.
- Architected a longitudinal feature-engineering framework over nearly **10 years of claims history**, processing **3–4M medical, pharmacy, genetics, and provider events** using PySpark and Azure Databricks and generating approximately **8,000 engineered patient features from 273 clinical feature groups**.
- Addressed severe data sparsity, noisy longitudinal histories, and substantial phenotypic overlap across LGMD subtypes; combined **133 genetically confirmed patients** with diagnosis-supported R9 patients while designing clinically anchored observation windows and subtype-aware feature families.
- Designed and evaluated supervised classification, Positive-Unlabeled learning, distance-based methods, dimensionality-reduction approaches, and rank-based model validation; achieved **55.2% precision within the top 1.5%** in TP-vs-RN cross-validation, representing approximately **5× enrichment over baseline**.
- Validated generalization into the unlabeled population using a controlled **20% spy-injection framework**, achieving approximately **15× lift over random patient selection** while preserving clinically expected R9 characteristics across cardiac, respiratory, muscular, mobility, and specialist-care features.
- Created the end-to-end DS lifecycle roadmap, decomposed delivery into BA and DS workstreams, reviewed analytical and modeling outputs, coordinated dependencies with clinical review, and consolidated findings into client-ready recommendations.
- Led recurring client working sessions, presenting all data science updates, defending cohort and modeling decisions, recommending subsequent iterations, and translating stakeholder and clinical feedback directly into analytical workstreams.

Associate - Data Scientist

ProcDNA Analytics

Jul 2025 – Jan 2026

Gurugram, India

HCP Look-Alike & Persona Modeling

- Built a scalable HCP look-alike modeling framework across a **150K+ HCP population** using XGBoost, LightGBM, Random Forest, and ensemble approaches for multi-class behavioral persona prediction.
- Addressed **70–80% feature sparsity** through targeted feature engineering, feature selection, and model ensembling, improving precision from approximately **25% to 65.34%** in the four-class classification setup.

- Translated model outputs into reusable HCP targeting cohorts, eliminating the need for repeated clustering-based persona rebuilds and improving scalability of downstream commercial targeting; solution received client recognition.

Enterprise Patient Event Prediction

- Productionized **two enterprise-scale batch inference pipelines** on Azure Databricks using PySpark, scoring approximately **1.2M patients across 800M+ longitudinal healthcare records**.
- Automated quarterly model scoring through reusable MLOps and CI/CD workflows, improving reliability and reducing manual intervention across recurring patient targeting cycles.
- Built scalable feature processing, model-scoring, and output-generation components designed for large healthcare datasets and downstream analytics consumption; initiative received project-level client recognition.

Data Scientist - Intern
ProcDNA Analytics

Jan 2025 – Jul 2025
Gurugram, India

Patient Discontinuation Prediction & Alert Generation

- Built an end-to-end patient discontinuation prediction system covering the full ML lifecycle from exploratory analysis and feature engineering through model validation and production deployment.
- Engineered scalable feature pipelines across **50M+ patient records** under constrained compute and memory environments and trained a Random Forest classifier achieving **75.3% ROC-AUC** with stable week-over-week performance of approximately $\pm 3\%$.
- Designed and deployed a containerized **alert-generation engine** using Docker and API-based integration to convert model predictions into actionable high-risk patient alerts.
- Independently owned the alert engine from design through deployment, earning the **Rising Star Award – Top 2 of 40 interns** for end-to-end project delivery.

SKILLS

Programming & Data	Python, SQL, PySpark, Pandas, NumPy, DuckDB, Git
Machine Learning	Scikit-learn, XGBoost, LightGBM, Random Forest, Ensemble Methods, Positive-Unlabeled Learning, Classification, Clustering, PCA, LDA, SHAP, Optuna
Deep Learning	PyTorch, TensorFlow
MLOps & Cloud	Azure Databricks, MLflow, Azure ML, Docker, Kubernetes, CI/CD, AWS SageMaker, AWS CodeBuild, APIs
Healthcare Analytics	Longitudinal Claims Analytics, Patient Identification, HCP Targeting, Clinical Feature Engineering, Rare-Disease Modeling

EDUCATION

B.E. Computer Science , Thapar University CGPA: 8.62/10	2021 – 2025
CBSE Class XII , BD Blue Bells Public School, Gurugram 92%	2021
CBSE Class X , BD Blue Bells Public School, Gurugram 89%	2019

AWARDS & ACHIEVEMENTS

- **Rising Star Award** – Ranked among the top 2 of 40 interns for end-to-end ownership and project delivery.
- Received **Client Recognition Award** for delivering high-impact data science solutions.
- Received project awards for **Patient Event Prediction** and **HCP Look-Alike Modeling** initiatives.
- Ranked **27th out of 400 teams** in IBM Datathon 2023.