

HARSHIT JAIN

(930)-904-1020 | jainharshitani@gmail.com | [linkedin.com/in/harshitaniijain](https://www.linkedin.com/in/harshitaniijain) | github.com/harshitjain0302

EXPERIENCE

Data Scientist

Jan 2026 – Present

Faculty Assistance in Data Science (FADS), Indiana University

Bloomington, IN

- Led end-to-end machine learning and analytics workflows to model chronological age and biological age acceleration using **10K+** facial images and large-scale speech datasets collected as part of a population-scale health study.
- Integrated multimodal data sources including facial imagery, acoustic feature representations, and epigenetic aging measures to benchmark model predictions against established biological aging clocks used in clinical research.
- Designed, executed, and analyzed model evaluation experiments, performing detailed error analysis, residual diagnostics, and subgroup performance comparisons to surface bias, instability, and failure conditions across demographics.
- Translated complex modeling results, assumptions, and limitations into clear, structured insights for non-technical stakeholders to inform interpretation, downstream analysis, and research decisions.

Research Data Scientist

Jun 2025 – Present

Data Science and AI Lab, Indiana University

Bloomington, IN

- Architected and maintained an end-to-end analytics pipeline ingesting metadata from **2.2M+ Hugging Face models**, enabling ecosystem-level visibility into open-source ML adoption, usage patterns, and security risk signals.
- Developed scalable SQL- and Python-based data processing workflows to extract, validate, and join tens of thousands of GitHub repositories, producing analytics-ready datasets used for large-scale exploratory and longitudinal analysis.
- Owned the definition, tracking, and interpretation of **10+ operational KPIs** including coverage rate, valid linkage rate, scan success rate, and failure recurrence to explain metric movement and guide iterative pipeline improvements.
- Executed vulnerability and asset analyses under the NSF Safe-OSE research grant, running automated security scans on Exosphere codebases and Jetstream2 images to identify systemic misconfigurations and recurring high-risk patterns.

Machine Learning Engineer (Analytics)

Jun 2023 – Aug 2023

Feynn Labs

Remote, India

- Analyzed **100K+ customer records** using SQL and Python to uncover behavioral trends, operational inefficiencies, and latent opportunities for improving customer engagement and product performance.
- Delivered reusable analytics pipelines and Power BI KPI reporting solutions to support recurring business reviews, reducing reporting turnaround time by **35%** and saving **10+ hours per week** for cross-functional teams.
- Drove data-backed recommendations for customer segmentation and product strategy by synthesizing analytical findings into clear, actionable insights for business stakeholders.

PROJECTS

Cardiovascular Health Analysis - SQL, BigQuery, Looker Studio

- Built a cloud-based analytics pipeline to preprocess and analyze **68K+ patient records**, engineering clinical and demographic features and delivering cohort-level dashboards that surfaced statistically significant cardiovascular risk drivers for stakeholders.

Neural Network-Based Image Analysis - Python, TensorFlow, Scikit-learn

- Trained and evaluated CNN models on **10K+ labeled images**, improving classification accuracy by **~25%** over baseline approaches and applying PCA and t-SNE to interpret embedding behavior and clustering structure.

SKILLS

Analytics & Statistics: Experiment Design (A/B, Quasi-Experiments), Regression, Classification, Clustering, Time-Series Analysis

Machine Learning: Feature Engineering, Tree-Based Models, Neural Networks, Anomaly Detection, Model Evaluation

Data & Programming: SQL, Python (Pandas, NumPy, SciPy, scikit-learn), BigQuery, Spark-style analytics, MLflow

Visualization & Reporting: KPI definition and monitoring, analytical dashboards (Tableau, Power BI), Matplotlib, Seaborn

EDUCATION

Indiana University Bloomington

Master of Science in Data Science

Aug 2024 – May 2026

Bloomington, IN

University of Mumbai

Bachelor of Technology in Electronics & Telecommunication Engineering

Dec 2020 – Jun 2024

Mumbai, India