

HARI ETTA

Data Scientist · AI Engineer

Chicago, IL · 312-721-3760

hetta.ds@gmail.com

[Github](#) · [LinkedIn](#)

SUMMARY

Data Scientist with an MS in Data Science & AI (IIT Chicago, 2026, GPA 3.6), driven by genuine curiosity for how machine learning solves real, high-stakes problems. Proven ability across predictive modeling, explainable AI, and production ML systems, comfortable questioning assumptions and digging into the "why" behind data. Skilled across the full ML lifecycle including Python, SQL, XGBoost, SHAP, and cloud deployment, with the ability to translate model behavior clearly for both technical and non-technical stakeholders.

TECHNICAL SKILLS

Analytics & BI	SQL · Tableau · Looker Studio · Power BI · A/B Testing · Cohort Analysis · Causal Inference · Statistical Testing
Languages	Python (pandas · NumPy · scikit-learn) · R · SQL
Data Engineering	dbt Core · Databricks · Delta Lake · Airflow · Prefect · Elasticsearch
Cloud & Platforms	GCP (BigQuery · Cloud Run · GCS) · AWS (S3 · Lambda) · Docker · FastAPI · GitHub Actions CI/CD
ML & AI	MLflow · XGBoost · SHAP · HuggingFace · RAG · LLMs · Azure OpenAI GPT-4.1

WORK EXPERIENCE

Associate — Design Technology & Client Analytics | Total Environment | Bengaluru, India July 2021 – July 2024

- Delivered 3–5 dashboards weekly for client meetings, internal reviews, and project status reporting, giving leadership real-time visibility into \$3–5M project portfolios.
- Managed end-to-end delivery of \$3–5M projects including laying out project plans, coordinating cross-functional teams, and dispatching instructions to architecture and engineering teams.
- Analyzed customer and operational data to identify bottlenecks, producing structured reports for senior leadership that reduced client deliverable turnaround time by 30%.
- Built internal performance dashboards consolidating operational metrics across departments, replacing manual multi-sheet tracking with a unified real-time reporting layer and translating findings into clear narratives for non-technical executives.

Data & Analytics Assistant | Chicago Public Schools at IIT Chicago | Chicago, IL Sep 2025 – Apr 2026

- Managed and maintained assessment records for 100–200 students daily across an 18-week semester, processing 3 exam cycles per student over **5,000+ individual data points** tracked in Excel.
- Performed end-of-day data entry and quality checks ensuring **100% accuracy** of student assessment records across all administrative systems.
- Generated structured analytics reports for faculty summarizing assessment trends, scheduling compliance, and cohort performance across student groups.

Data Scientist & VC Analyst | UruQAI — Quantum Computing Startup | Uruguay (Remote) Jan 2026 – Apr 2026

- Built predictive scoring models and analytics pipelines on early-stage investment data, applying quantitative traction analysis to evaluate company performance contributing to a 25x increase in investor engagement from 2 to 50 interactions over 8 weeks.
- Designed Python-based analysis pipelines to process quantum experiment results, identifying performance patterns and error rates in circuit behavior that directly informed algorithm optimization and model improvement decisions.
- Translated complex technical and statistical findings into clear, decision-ready narratives for engineering teams and investors, bridging experimental research with applied modeling and strategic outcomes.

ANALYTICS PROJECTS

Startup Funding Intelligence & Market Dashboard — [Repository](#)

Stack: Databricks · Spark NLP · dbt Core · MLflow · HuggingFace · ChromaDB · Looker Studio · FastAPI · GCP

- Built an NLP-based sector classification model on Databricks using Apache Spark, automating startup pattern recognition and reducing manual deal-triage time by an estimated 70%.
- Designed an end-to-end ML pipeline with dbt transformations and LLM-based feature extraction, generating a data-driven VC investment thesis that identified underserved market sectors.
- Enforced Pydantic schema validation on all LLM outputs before downstream consumption, ensuring model reliability and preventing corrupted predictions from propagating through the analytics pipeline.

Patient Readmission Analytics & Prediction System — [Repository](#)

Stack: Python · XGBoost · SHAP · MLflow · Streamlit · FastAPI · Docker · GCP Cloud Run · GitHub Actions

- Built an ML pipeline to predict 30-day hospital readmission risk from 70K diabetic patient encounters, aiming to help flag high-risk patients before discharge.
- Handled severe 11:89 class imbalance using SMOTE and cost-sensitive threshold tuning rather than optimizing for accuracy, which would have been misleading on this dataset.
- Benchmarked Logistic Regression, Random Forest, and XGBoost with MLflow experiment tracking; selected the final model using a custom clinical cost function, achieving 0.65 AUC-ROC and 99% recall on high-risk patients.

IIT International Student Policy Chatbot — [Repository](#)

Stack: Python · FastAPI · Azure OpenAI GPT-4.1 · Elasticsearch · RAG · Next.js · Neon PostgreSQL · Netlify

- Achieved **90.4% accuracy** on a 52-question benchmark, outperforming AutoRAG and NotebookLM baselines deployed live 24/7 with real IIT students in closed beta.
- Designed a Neon PostgreSQL feedback analytics layer capturing every user session and query, enabling continuous performance monitoring and iterative improvement post-launch.
- Built hybrid BM25 + vector search with RRF fusion across 328 policy document chunks, improving retrieval precision across both keyword and semantic query types.

EDUCATION

Illinois Institute of Technology, Chicago

MS — Data Science & Applied AI (Information Technology & Management) · GPA: 3.6 / 4.0

Aug 2024 – May 2026

Manipal Academy of Higher Education, India

Bachelor of Engineering · GPA: 3.1 / 4.0

July 2017 – July 2021