Hari Sreedeth

Applied Data Scientist | ML Engineering & Deployment

Address: 1 Queen St, Randwick NSW 2031 | Phone: 0406058378 |
e-mail: harisreedeth001@gmail.com
LinkedIn: https://www.linkedin.com/in/harisreedeth/
GitHub: https://github.com/hsreedeth | Portfolio Page: https://hsreedeth.github.io/portfolio/

Professional Summary

Data Science postgraduate with a strong foundation in Python, SQL, and building reproducible machine learning pipelines. Experienced in engineering end-to-end ML workflows and processing complex, large-scale datasets (500,000+ records) using strict version control (Git). Highly motivated to transition into applied ML engineering, with a strong interest in model lifecycle management, API integration, and cloud platforms (GCP). A collaborative communicator eager to learn best practices for deploying data science models into production environments.

Core Skills

Programming and Databases

Languages: Python (pandas, NumPy, scikit-learn), SQL, Bash/shell scripting, R.
Data Processing: Large-scale data manipulation, complex SQL querying, dataset linkage, and rigorous quality control.

Machine Learning & Engineering Pipelines

ML Workflows: End-to-end pipeline development, feature engineering, predictive modelling, and unsupervised clustering.
Tools & Frameworks: Jupyter environments, workflow managers (Snakemake, Nextflow), and reproducible analysis patterns.
Deployment Mindset: Working knowledge of integrating models via APIs and a strong interest in cloud platforms (GCP/AWS).

Infrastructure & Research Computing

Version Control: Advanced Git/GitHub for collaborative and auditable codebases.
Environment: Unix/Linux command line proficiency.
Compute: Experience executing and monitoring resource-intensive jobs on SLURM-based HPC and remote clusters.

Collaboration & Delivery

Communication: Strong ability to translate complex statistical outputs into actionable insights for diverse stakeholders.
Agile/Team Fit: Collaborative mindset with a proven history of working cross-functionally to solve technical and domain-specific problems.

Experience

Postgraduate Researcher (Data Science Workflows) | The George Institute for Global Health, Sydney
2025 – Present

- **Engineered Data Pipelines**: Architected an end-to-end Python and SQL data processing pipeline to extract, clean, and harmonize complex, high-dimensional records for over 500,000 participants.
- **Built ML Workflows**: Developed reproducible and auditable machine learning workflows in secure remote and HPC (High-Performance Computing) environments, utilizing strict Git version control to manage the codebase.
- **Cross-Functional Collaboration**: Collaborated closely with senior researchers and domain experts to translate complex statistical and ML outputs into actionable insights, data visualizations, and structured reports.

Clinical & Research Pharmacist | GMG Healthcare, Dubai (UAE)
2020 – 2023

- **Data Extraction & Analytics**: Supported institutional research by executing structured data extraction, rigorous data cleaning, and providing analytical support for observational studies.
- **Stakeholder Communication**: Operated in a fast-paced, multidisciplinary environment, regularly translating highly technical and clinical information for diverse stakeholders.
- **Process Optimization**: Maintained rigorous attention to detail, quality control, and compliance while managing high-throughput clinical data and systems.

Selected Research & Data Science Projects

Clinical Trials Data Platform & ML Pipeline | Research Organization (Team Project)

- **Data Warehouse Architecture:** Engineered a production-grade data platform using Docker and Postgres to automatically ingest and store massive relational data snapshots from ClinicalTrials.gov.
- **ELT & Data Quality Gates:** Architected an end-to-end transformation pipeline using dbt (data build tool) to build query-ready dimensional marts (star schema). Implemented strict data-quality testing and integrated them into a GitHub Actions CI pipeline to prevent broken models from merging.
- **Applied ML & Deployment:** Collaborated with a team of engineers to deploy a Streamlit dashboard serving real-time analytics, alongside integrating a lightweight predictive machine learning model to forecast trial duration and early termination risk.

UK Biobank: Large-Scale ML Pipeline | The George Institute / UNSW

- **High-Volume Data Processing:** Engineered a scalable Python and SQL data pipeline to extract, clean, and model complex records for a massive cohort of over 500,000 participants.
- **Infrastructure & Version Control:** Executed resource-intensive machine learning and statistical workflows on remote SLURM-based HPC clusters. Strictly utilized Git version control to guarantee a fully reproducible and auditable codebase for production handover.

MAIP: Applied ML & LLM Integration | UNSW

- **Unsupervised ML Pipeline:** Built a robust clustering and phenotyping pipeline to process highly complex, high-dimensional datasets into stable, distinct subgroups.
- **LLM Rule-Translation Layer:** Engineered a surrogate translation layer leveraging a Large Language Model (LLM) to automatically convert complex machine learning outputs into transparent, readable decision rules for non-technical end-users.

Publications & Presentations

- Co-author, Transforming Strategy with AI-Augmented Personas for Sustainability, Data Management, Regulatory Frameworks, Healthcare and Compliance, ADIPEC 2024 (healthcare use-case design). (DOI: 10.2118/222041-MS)
- Co-author, Opinion of Community Pharmacists in Ajman on the International Pharmaceutical Federation's Workforce Development Goals, presented at ADPHAC 2019 and the 79th FIP World Congress.(https://shorturl.at/R2j0t)
- Sex-Specific Multimorbidity Clusters Using First Occurrence Data from the UK Biobank – MSc dissertation. Manuscript in preparation.

Education

University of New South Wales, Sydney
Master of Science in Health Data Science (Extension)
2024 – 2025

Core: Statistical Modelling, Machine Learning in Health Data Analytics, Management & Curation of Health Data, Computation for Health Data Science
Electives: Applied Bioinformatics, Database Systems
Graduated with a distinction average grade of 80 and High Distinction in Master's Dissertation (87).

Gulf Medical University, UAE
Doctor of Pharmacy
2014 - 2019

Undergraduate program with major focus on Clinical Pharmacy. Graduated with a 3.65 GPA.