

Vineeth Adapa

(470)-286-0422 | vineethadapa01@gmail.com | Washington, DC | LinkedIn | GitHub | Portfolio

EXPERIENCE

Data Science Intern

Sep 2025 - Present

Washington, DC

Pan American Health Organization / World Health Organization

- Develop multilingual R pipelines to clean, deduplicate, and integrate over 70,000 research records, creating a unified bibliometric dataset for PAJPH.
- Architect LLM-powered NLP workflows with Llama 4 Scout to extract and classify research keywords into WHO aligned public health domains.
- Generate semantic embeddings from abstracts, keywords, and titles using OpenAI's text-embedding-3-large model to enable article similarity recommendations and related article discovery.
- Implement production grade LLMops including token-aware batching, automated JSON repair, and adaptive rate limiting to support reliable, large-scale inference.
- Partner with editorial and bibliometrics teams to collect, validate, and standardize datasets enabling journal impact factor calculations and performance analysis.
- Design A/B testing infrastructure to measure editorial policies and citation interventions, working with PAJPH regional advisors to inform evidence based improvements.

Data Science Volunteer

Feb 2025 - Sep 2025

Seattle, WA

Thriving Elements

- Designed and deployed scalable mentor-mentee recommendation system using TF-IDF and cosine similarity, cutting manual matching effort by 14% across cohorts.
- Improved pairing relevance by combining sentence-BERT embeddings with FAISS approximate nearest neighbor search, increasing post match satisfaction scores by 12% per survey.
- Engineered features from participants profiles (skills, goals, availability, engagement history) in Python to feed clean inputs into recsys pipeline.
- Assessed performance via Precision@3, NDCG, and A/B test, delivering an 8% lift in matching accuracy over baseline heuristics.
- Collaborated with program managers to iterate on matching logic, incorporate user feedback, and align systems with organizational priorities.

Data Analyst

Feb 2021 - Dec 2022

Hyderabad, India

Cognizant Technology Solutions

- Built and maintained data ingestion, transformation, and validation pipelines in SQL Server and Excel to ensure high-quality client datasets for financial services engagements.
- Optimized SQL queries, stored procedures, and reporting scripts, reducing analytical turnaround time and improving deliverable reliability.
- Developed interactive Power BI dashboards visualizing key KPIs and business metrics to drive stakeholder decision-making on operations and finance.
- Performed data quality audits, root cause analysis of issues, and documented resolutions, collaborated in Agile sprints to test and deploy production analytics solutions.

SKILLS

Programming Languages:

Python (Pandas, NumPy), R, SQL (SQL Server)

Machine Learning:

Scikit-learn, PyTorch, Causal ML, Uplift Modeling, Recommendation Systems

Gen AI & NLP:

LLMs(Llama 4, GPT), RAG, HuggingFace, BERT, FAISS, spaCy, NLTK

Data Engineering:

Apache Spark, Airflow, Snowflake, Data Ingestion & ETL Pipelines

MLOps & Deployment:

MLflow, Docker, Flask, FastAPI, LLMops

Stats & Experimentation:

A/B Testing, Hypothesis Testing, Bootstrapping, Causal Inference

Cloud & Tools:

AWS, Git, Agile Methodologies

PROJECTS

Uplift Modeling & A/B Testing

Built a causal ML system estimating uplift from promotional banners, enabling targeted incentives that increase incremental conversions and reduce wasted marketing spend.

Causal Simulation of Vaccination Policies

Designing a stochastic, risk-stratified SEIR simulation with counterfactual causal inference to quantify the mortality and system load impact of optimized vaccination policies versus real world heuristics.

EDUCATION

Master's of Science in Information Systems

Jan 2023 - Dec 2024

Focus Area: Data Mining, Statistical Data Analysis, Advanced Databases