Vineeth Adapa

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

EXPERIENCE

Data Science Intern **Sep 2025 - Present**

Pan American Health Organization / World Health Organization Washington, DC

- Build multilingual R pipelines to clean, deduplicate, and integrate over 70,000 research records, standardizing text-heavy abstracts and metadata into a unified bibliometric dataset for PAJPH.
- Design and implement LLM powered NLP workflows with Llama-based models to normalize input text, generate prompts, and classify research keywords into WHO aligned public health domains, in collaboration with data scientists.
- Generate semantic embeddings from abstracts, keywords, and titles using OpenAI's text-embedding-3-large model to support article similarity search and related article recommendations for researchers.
- Contribute to internal LLM workflows by helping configure token aware batching, JSON output repair, and rate limiting strategies, work alongside senior team members to enable reliable large scale inference in a research setting.
- Partner with editorial and bibliometrics teams to collect, validate, and standardize datasets enabling journal impact factor calculations and performance analysis.
- Drafted an experimental framework for A/B testing editorial policies and citation interventions, defining key metrics and proposed designs to support evidence based improvements once implemented.

Data Science Volunteer **Feb 2025 - Sep 2025**

Thriving Elements Seattle, WA

- Designed and deployed a mentor-mentee recommendation prototype using TF-IDF and cosine similarity to reduce manual effort in matching cohorts.
- Improved pairing relevance by incorporating sentence-BERT embeddings with FAISS approximate nearest neighbor search, and used follow-up surveys to capture satisfaction with the recommended matches.
- Engineered features from participants profiles (skills, goals, availability, engagement history) in Python to create structured inputs for the recommendation pipeline.
- Evaluated recommendation quality with offline metrics such as Precision@3 and NDCG, manually reviewing recommendations on historical data against heuristic baselines.
- Collaborated with program managers to refine matching logic, incorporate qualitative feedback from mentors and mentees, and align the system with organizational priorities.

Data Analyst **Feb 2021 - Dec 2022**

Cognizant Technology Solutions Hyderabad, India

- Built and maintained data ingestion, transformation, and validation workflows in SQL Server and Excel to ensure high-quality datasets for financial services clients.
- Optimized SQL queries, stored procedures, and reporting scripts, reducing turnaround time for recurring analyses and improving reliability of client deliverables.
- Developed interactive Power BI dashboards to visualize key KPIs and business metrics, enabling stakeholders to monitor operations and financial performance.
- 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 family, GPT), RAG concepts, HuggingFace, BERT, FAISS, spaCy, NLTK
Data & Pipelines:	Data Ingestion and ETL pipelines, basic batch processing
MLOps & Deployment:	MLflow, Docker, Flask, FastAPI, LLMops
Stats & Experimentation:	A/B Testing concepts, Hypothesis Testing, Bootstrapping, Causal Inference
Cloud & Tools:	AWS, Git, Agile Methodologies

RESEARCH

Uplift Modeling & A/B Testing

- Building a T-learner uplift modeling to estimate incremental impact of promotional banners on an open-source marketing dataset.
- Simulates treatment/control assignment and evaluates uplift models to study how model choice and features influence accuracy.

Causal Simulation of Vaccination Policies

- Developing a stochastic, risk-stratified SEIR simulation in Python to assess mortality and system load impacts of alternative strategies.
- Calibrates model parameters to synthetic population data and uses bootstrapping to quantify uncertainty across policy scenarios.

EDUCATION

Master's of Science in Information Systems **Jan 2023 - Dec 2024**

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