

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

- Build R pipelines to clean, deduplicate, and integrate 70,000+ multilingual research records, creating a unified bibliometric dataset used for journal analytics.
- Design NLP workflows using Llama-based models to normalize text, classify research topics, and align metadata with WHO public health domains.
- Generate semantic embeddings using OpenAI text-embedding-3-large to support article similarity search and related article recommendations.
- Improve internal LLM workflows by configuring token-aware batching, JSON output repair, and rate-limiting strategies for reliable large scale inference.
- Collaborate with editorial and bibliometrics teams to validate datasets supporting impact factor calculations and journal performance analysis.
- Draft an experimental framework for A/B testing editorial policies, defining metrics and propose designs to evaluate citation related interventions.

Data Science Volunteer

Feb 2025 - Sep 2025

Seattle, WA

Thriving Elements

- Cleaned and standardized mentor/mentee profile data (skills, interests, goals, availability) to create structured inputs for matching.
- Built a TF-IDF + cosine similarity prototype to reduce manual pairing effort for program managers.
- Analyzed historical pairing decisions and satisfaction surveys to identify patterns that improved matching logic.
- Collaborated with program staff to refine rules and incorporate qualitative feedback into the recommendation process.

Data Analyst

Feb 2021 - Dec 2022

Hyderabad, India

Cognizant Technology Solutions

- Built and maintained SQL Server pipelines for data ingestion, transformation, and validation, ensuring high quality datasets for financial services clients.
- Optimized SQL queries and stored procedures, reducing turnaround time for recurring analyses.
- Developed Power BI dashboards to visualize KPIs and operational metrics for business stakeholders.
- Performed data quality audits, root cause analysis, and collaborated in Agile sprints to deploy production analytics solutions.

PROJECTS

Causal Simulation of Vaccination Policies

- Developed an agent based SEIR epidemic simulator with individual level feature modeling to estimate vaccine allocation policies under supply constraints.
- Implemented propensity score estimation to address confounding in treatment assignment for causal off-policy evaluation.

Uplift Modeling & A/B Testing for E-commerce

- Built end-to-end causal inference pipeline for promotional targeting, engineered RFM features from 540k transactions, validated experimental design via z-test, and T-leaner to estimate individual treatment effects across 4k customers in 10 uplift deciles.
- Applied Uplift modeling to optimize promotional spend, analysis revealed 40% of customers as high-value targets while identifying 30% as sleeping dogs, where promotional treatment reduced conversion rates by up to 7.5%.

Air Quality Forecasting

- Built a multi-model forecasting pipeline in R to predict Temperature, O₃ and PM2.5 levels using time series models with early stopping, achieving RMSE reductions of 12-18% across pollutants after feature engineering with 5-day lag variables.
- Computed EPA standard AQI from predicted pollutant levels and developed an interactive Shiny dashboard with geospatial mapping, enabling 10-day AQI forecasts across U.S. locations and supporting trend analysis.

SKILLS

Programming Languages: Python (Pandas, NumPy), R, SQL (SQL Server)

Machine Learning: Supervised ML, Scikit-learn, Feature Engineering, Model Evaluation, Uplift modeling, Recommender Systems

Gen AI & NLP: LLMs (Llama family, GPT), RAG concepts, HuggingFace, BERT, spaCy, NLTK

Data & Pipelines: ETL, Data Cleaning, Batch Processing, Power BI

MLOps: MLflow, Docker, Flask, FastAPI, LLMops fundamentals

Stats & Experimentation: A/B Testing, Hypothesis Testing, Bootstrapping, Casual Inference

Cloud & Tools: AWS, Git, Agile

EDUCATION

Master of Science in Information Systems

Jan 2023 - Dec 2024

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