

# Vineeth Adapa

Data scientist with strong applied statistics expertise and end-to-end experience in modeling and experimentation

vineethadapa01@gmail.com  
(470) 286-0422  
github.com/Vineeth-chaitanya  
Portfolio/Vineeth-chaitanya

## Relevant Experience

### Data Science Intern • Pan American Health Organization/ WHO

Sept 2025 - Present

- Built and executed multilingual R pipelines to clean, and integrate over 70,000 research records, creating a unified bibliometric dataset for journal performance analysis.
- Developed classification models to categorize publications into public health domains, achieving 80% agreement with expert manual labels on a 1000 record validation set.
- Implemented scalable similarity analysis and retrieval mechanisms to enable fast related article discovery and internal search tools with sub second response times.
- Optimized model inference workflows through batch processing, output validation, throughput enhancements, increasing processing capacity by 2 times.
- Partnered with editorial and analytics teams to standardize datasets for performance reporting and design experimental frameworks (A/B Testing) to evaluate policies.

### Data Engineer Intern • Maxpep LLC

Feb 2025 - Sept 2025

- Engineered Python based ingestion pipelines for 120,000 profile and text records from diverse sources, reducing preprocessing runtime from 45 to 18 minutes.
- Designed feature engineering pipelines combining statistical (TF-IDF) and advanced representation techniques, boosting top-3 match precision by 17%.
- Deployed efficient approximate nearest-neighbor indexing for similarity search, cutting average query latency from 2.1 seconds to 180 ms in prototype systems.
- Introduced data quality controls (schema validation, outlier detection) that decreased pipeline errors by 30% during iterative weekly experiments.
- Created an offline evaluation framework using Precision@K and NDCG metrics on a 500-pair ground truth dataset to benchmark and improve matching performance.

### Data Analyst • Cognizant Technology Solutions

Feb 2021 - Dec 2022

- Developed and optimized SQL Server pipelines for financial services reporting, reducing recurring analysis turnaround by 30% and enhancing data accuracy.
- Built Power BI dashboards with embedded quality check to monitor KPIs, enabling rapid detection and operational visibility for stakeholders.

## Selected Projects

### Uplift Modeling & A/B Testing

Developed an end-to-end uplift modeling pipeline on 540,000 transactions using RFM features and T-learner approach; validated experimental setup with statistical test and identified high-value segments(40% uplift potential) alongside “sleeping dogs” where intervention reduced conversion by up to 75%.

### Causal Simulation of Vaccination Policies

Built a feature rich, 50000-agent SEIR simulation modeling California demographics under confounded vaccine allocation policies; conducted off-policy evaluation demonstrating 65.5% mortality reduction potential with optimized allocation strategies.

## Skills

### Programming Languages

R, Python, SQL

### Machine learning

Scikit-learn, PyTorch, Causal ML / Causal Inference, Uplift Modeling, Recommendation systems

### Statistics & Experimentation

A/B Testing, Hypothesis Testing, Bootstrapping, Statistical Analysis

### MLOps & Tools

Docker, FastAPI, Git, AWS, Data Pipelines /ETL

## Education

### University of Maryland-Baltimore

Jan 2023 - Dec 2024

Masters of Science in Information Systems  
Data Science, Advanced Databases, Statistical Analysis

## Volunteer

### Thriving elements

Designed and maintained program level performance metrics, analyzing mentor-mentee engagement, match quality, and retention trends to guide leadership decisions and improve operational effectiveness.