

DharmpratapSingh Vaghela

Chantilly, Virginia | (571) 349-6301 | dharmpratapv@gmail.com | [LinkedIn](#) | [GitHub](#)

Graduate student in Data Analytics Engineering with hands-on AWS production experience and expertise in large-scale ML pipelines. Currently building financial integrity systems on AWS (EC2, S3) and developing ClimateGPT, a production LLM integration for climate data analytics. Specialized in cloud-native data solutions and real-time analytics.

EDUCATION

George Mason University, Fairfax, VA, USA Jan 2024 - Expected Graduation Dec 2025

Master of Science in Data Analytics Engineering GPA: 3.67

- Relevant Coursework: Applied Predictive Analytics, Interpretable Machine Learning, Bayesian Artificial Intelligence, Data Management and Mining, Visualization of Data.

Dharmsinh Desai University, Nadiad, Gujarat, India July 2018 – May 2023

Bachelor of Technology in Information Technology GPA: 3.25

Certifications: [Google Advanced Data Analytics Certification](#) and [AWS Data Engineer Badge](#).

PROJECTS

ClimateGPT LLM and Data Engineering Capstone Aug 2025 - Present

Technologies: Python, DuckDB, Docker, Streamlit, Git, FAIR metadata, Anthropic's MCP

- Designed and containerized a Model Context Protocol (MCP) server for ClimateGPT, enabling real-time, natural-language querying of large-scale emissions data.
- Processed, curated, and normalized 7GB+ of CO₂ emission data from 2000 to 2023 for data-driven climate analysis at various levels.
- Reduced dataset storage by 95% (from ~7GB to 300MB) and automated ingestion for real-time LLM-based querying.
- Enabled complete reproducibility and improved team efficiency by containerizing the MCP server and deploying scalable pipelines for capstone demo and Erasmus AI partners.

Reddit Sentiment Analysis Dashboard | [GitHub Link](#) Dec 2024 – March 2025

Technologies: Python, Streamlit, PRAW, HuggingFace Transformers, Plotly, Pandas, PyTorch

- Engineered a real-time, multi-model NLP dashboard using four sentiment analysis models (Cardiff, DistilBERT, NLPTown, BERTweet) to evaluate Reddit posts.
- Developed dynamic visualizations (temporal trends, bar and radar charts) for rapid insights.
- Integrated PRAW for scalable Reddit data ingestion and generated Streamlit UI for better understanding of insights.

Walacor Financial Integrity Sept 2025 – Present

Technologies: Python, AWS EC2, S3, Next JS, Rest API, Walacor Blockchain

- Currently architecting and deploying a cloud-native financial data integrity platform on AWS, utilizing EC2 for compute and S3 for scalable data storage
- Implementing fraud detection, anomaly detection and compliance monitoring using 256-bit hashing algorithms to secure data, applying AWS best practices for security and cost optimization.
- Gaining hands-on production experience with AWS infrastructure, IAM policies, and cloud-native architecture for financial systems.

EXPERIENCE

Data Analytics Intern - Atharva Systems, Ahmedabad, Gujarat December 2022 – June 2023

- Automated analysis of large ERP datasets using Python and Excel, delivering actionable retention and sales insights for 5+ cross-functional teams.
- Designed interactive PowerBI dashboards for real-time KPI reporting, reducing manual reporting time by 30% and enabling data-driven decision-making at the manager level.
- Optimized SQL queries for ODOO/PostgreSQL databases, decreasing average data retrieval and processing time by 20%, improving report latency for analytics users.

TECHNICAL SKILLS

- **Programming:** Python (NumPy, Pandas, scikit-learn, TensorFlow, PyTorch), SQL, R, C++, Java
- **Machine Learning:** Regression, Classification, Random Forest, XGBoost, NLP, LSTM, Deep Learning
- **Data Engineering:** ETL Pipelines, Docker, Airflow, Databricks, Spark, DuckDB, AWS, Azure, GCP
- **Visualization:** Power BI, Tableau, Matplotlib, Seaborn, Plotly
- **Cloud/Database:** PostgreSQL, MongoDB, AWS Redshift, S3, Glue, Lambda, EC2
- **Other:** Agile, MLOps, CI/CD, FAIR data, Git