DharmpratapSingh Vaghela

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

Data Analytics Engineer with expertise in LLM integration (MCP, RAG), real-time NLP pipelines, and AWS cloud infrastructure. Built ClimateGPT's Model Context Protocol server, multi-model sentiment analysis dashboards, and a hackathon-winning blockchain fraud detection platform. Seeking Data Scientist roles in fintech or climate analytics.

EDUCATION

George Mason University, Fairfax, VA, USA   Jan 2024 - Dec 2025
Master of Science in Data Analytics Engineering   GPA: 3.77
- 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 – Dec 2025
Technologies: Python, DuckDB, Docker, Streamlit, Git, FAIR metadata, Anthropic's MCP
- Served as Lead Developer coordinating 6-person team while designing and containerizing 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 CO_2 emission data from 2000 to 2023 for data-driven climate analysis at various levels.
- Developed automated testing frameworks reducing validation cycles from 45 minutes to 3 minutes and created baseline knowledge guardrails that reduced AI hallucinations by 80%.

IntegrityX | ChallengeX Hackathon – 1st Place   Sept 2025 – Dec 2025
Technologies: Python, AWS EC2, S3, Next.js, REST API, Walacor Blockchain
- Won first place building a blockchain-based financial integrity platform for fraud detection and compliance monitoring in a 24-hour hackathon.
- Implemented anomaly detection using 256-bit hashing algorithms to secure financial transactions, applying AWS best practices for security and cost optimization.

Reddit Sentiment Analysis Dashboard   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.

EXPERIENCE

ODOO 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