

HARSH MALVIYA

Data Scientist • AI Engineer • Geospatial & Network Analytics Specialist
+1 (332) 207-5607 • hmalviya1@umassd.edu • LinkedIn • GitHub • harshmalviya.com
Dartmouth, MA • F-1 OPT / STEM OPT — up to 3 years work authorization

PROFESSIONAL SUMMARY

Analytical Data Scientist and AI Engineer with 2.5+ years of experience building machine learning pipelines, RAG systems, and geospatial network models. Published researcher in earthquake pattern analysis and global supply chain network science. Proficient in Python, LangChain, LLMs, SQL, and cloud-ready deployment (Docker, Streamlit). Seeking Data Scientist, ML Engineer, or AI Engineer roles where deep research skills and production-grade AI experience create real-world impact.

TECHNICAL SKILLS

Languages & Databases: Python, SQL, R, C/C++, MATLAB, MySQL, PostgreSQL, Oracle

AI / ML / NLP: LangChain, LLMs, RAG (FAISS + BM25 hybrid), Predictive Modeling, Ensemble Methods, Clustering, Time-Series Forecasting, NLP

Geospatial & Networks: GeoPandas, NetworkX, Small-World Network Analysis, Geospatial Modeling

Data & BI Tools: Power BI, Tableau, Qlik, Excel, Hadoop, ETL

Deployment & DevOps: Docker, Streamlit, Git, Linux, SLURM / HPC (Carnie Supercomputer)

Testing & Automation: Selenium IDE, Workflow Automation, System Administration

WORK EXPERIENCE

AI Engineer Intern | **Gogentic AI** | Texas, USA (Remote)

Jun 2025 – Aug 2025

- Designed and deployed production-grade Retrieval-Augmented Generation (RAG) pipelines using Python, LangChain, FAISS, and BM25 hybrid retrieval for enhanced analytics on Neurovault datasets
- Built a Supply Chain Risk Copilot with Streamlit frontend and Docker deployment, enabling real-time risk querying across structured and unstructured data sources
- Developed AI-powered meeting summarization tools with real-time voice-to-text transcription and privacy-aware analytics
- Integrated Oracle and PostgreSQL data sources into automated model workflows, improving intelligence pipeline efficiency

Web Application Architect | **Stadium Span** | Madhya Pradesh, India

Jul 2023 – Jul 2024

- Designed and developed SQL-based reports and interactive dashboards in Power BI and Tableau, improving operational efficiency by 20%
- Automated data scrubbing and validation pipelines using Python and SQL, improving system accuracy by 15%
- Built predictive analytics tools and integrated multi-source data pipelines aligned with business intelligence requirements
- Collaborated cross-functionally to implement data governance standards and authored detailed process documentation

Web Application Architect Intern | **Stadium Span** | Madhya Pradesh, India

Jan 2023 – Jun 2023

- Implemented data integration solutions and built early-stage predictive analytics tools using Python and SQL
- Leveraged Power BI to visualize complex datasets and deliver actionable trade data insights recognized by stakeholders

EDUCATION

Master of Science — Data Science
University of Massachusetts Dartmouth, Dartmouth, MA

2024 – Present

Relevant: High-Performance Computing, Advanced Statistics, Small World Networks, Advanced Data Mining, Business Intelligence, Database Design, Operating Systems

Bachelor of Technology — Computer Science Engineering

2019 – 2023

Rajiv Gandhi Proudlyogiki Vishwavidyalaya, Bhopal, India

Relevant: Machine Learning, Data Mining, Cloud Computing, Database Management, Computer Networks, IoT

KEY PROJECTS

Global Trade Network Analysis (MS Thesis)

- Modeled 35 years of UN Comtrade data (238 countries, 16,000+ links, 1988–2022) as geospatially embedded small-world networks using Python, NetworkX, and GeoPandas
- Confirmed 8/9 hypotheses including 80% supply chain fragmentation from China hub removal and Asia-Pacific clustering coefficient $C = 0.91$
- Ran large-scale simulations on Carnie HPC cluster; manuscript in submission to peer-reviewed journal

GitHub: github.com/harsh-stack/TradeNetworkAnalysis

Supply Chain Risk Copilot (RAG System)

- Built a production-grade AI copilot using LangChain, hybrid FAISS + BM25 retrieval, Streamlit UI, and Docker — enabling natural language risk querying over supply chain datasets
- Integrated multiple data sources including structured SQL and unstructured documents with semantic and keyword search fusion

Earthquake Pattern Analysis

- Applied K-Means clustering and ARIMA time-series forecasting to 27,696 global earthquake records (1960–2023), achieving ~87% accuracy in high-risk zone identification
- Published findings in peer-reviewed journal (QUEST Journals, 2025); interactive geospatial heatmap deployed as production-ready visualization

GitHub: github.com/harsh-stack/Earthquake-Pattern-Analysis

RESEARCH & PUBLICATIONS

Earthquake Pattern Analysis Using Clustering, Forecasting & Machine Learning

QUEST Journals — Journal of Research in Environmental and Earth Sciences | Published 2025

Global study across 100+ geographic regions using ensemble ML methods. 90%+ precision on critical seismic vulnerability zones. [\[ResearchGate\]](#)

Small-World Spatial Network Analysis of Global Supply Chains (1988–2023)

Manuscript in submission — Target: Peer-Reviewed Journal (Supply Chain / Network Science)

Confirmed small-world properties, power-law degree distributions, and asymmetric resilience under targeted hub disruptions in global trade networks.