

HARSH VERMA

✉ hverma4@pride.hofstra.edu  linkedin.com/in/veharsh  github.com/Vermahash 🏠 New York, USA

PROFESSIONAL SUMMARY

Professional with nearly three years of experience at Deloitte and Altice USA delivering high-impact data transformations, executive reporting, and predictive analytics for Fortune 500 clients. Expert at converting complex datasets into actionable C-suite insights via advanced dashboards, ML models, and automated pipelines. Continuously expanding consulting expertise through the **McKinsey Forward Program** and driven to fuel scalable impact through innovation and collaboration.

TECHNICAL SKILLS

Core Competencies: Data Analysis, Predictive Modeling, Machine Learning (ANN, Random Forest, Ensemble), Unsupervised Learning, Web Scraping, NLP, RAG, Summarization, Sentiment Analysis, Data Visualization, ETL, Dashboarding, Stakeholder Engagement, CLI Automation

Programming & Tools: Python, SQL, R, Git, Jira, Jupyter, VS Code, FastAPI, NGINX

Libraries & Frameworks: Pandas, NumPy, Scikit-learn, PyTorch, NLTK, LangChain, sentence-transformers, Neuralnet (R)

Visualization & Reporting: Tableau, Power BI, Data Storytelling, Advanced Dashboarding

Cloud & Deployment: AWS (Lambda, EC2, S3), API Development, Serverless Computing, Docker, Streamlit, MinIO, HuggingFace Hub

Databases & ETL: MySQL, PostgreSQL, DynamoDB, SQLAlchemy, Alembic, Workday, Pinecone

AI/ML Ops & Vector Stores: MLflow, Weights & Biases, FAISS, Chroma

EXPERIENCE

Optimum (Altice USA)

Feb 2025 – May 2025

Data Analytics, Strategic Program Office

Long Island City, NY

- Created automated executive dashboards in Power BI, powered by SQL and Python ETL pipelines, reducing reporting time by 70% and delivering real time KPI visibility to the CEO and SVP.
- Proposed a predictive analytics framework to forecast ROI risk and benefit realization early in the project lifecycle, enabling proactive decisions across service channels and maintaining on-target performance.
- Built VBA scripts that harmonized operational metrics with FP&A financial data, mapping cost consumption and projecting ROI for upcoming initiatives.
- Collaborated with cross functional teams to define KPIs and embedded them in Power BI dashboards, enabling stakeholders to visualize performance and enact informed policy changes
- Deployed automated Power BI data models and refresh routines, providing real-time dashboards that drove transparency and trust

Deloitte (Offices of the US)

May 2021 – Aug 2023

Data Analyst (Workday Practice)

India

- Migrated over 1M+ records for Fortune 500 clients (AT&T, Wells Fargo) to Workday using custom ETL pipelines with 98% accuracy.
- Automated ETL workflows with 100+ SQL scripts, cutting processing time by 40% & data mismatches(received vs intended) by 25%.
- Served as SME for client data resolution, deploying error-detection scripts that cut support tickets & reduced data error rates by 30%.
- Created monitoring dashboards to improve defect tracking and reduce manual effort by 50%.
- Led team onboarding efforts by launching an internal knowledge-sharing hub for data processes.

PROJECTS

NLP Automation & Summarization

- **News Scraping & LLM-Driven Summarization:** Built a Python pipeline with Selenium and BeautifulSoup to scrape NPR articles, generated on-device summaries using Facebook BART (HuggingFace), applied NLTK VADER for sentiment analysis, and delivered concise bullet-point reports with “Read More” links via a CLI.
- **NLP & Retrieval-Augmented Generation Applications:** Medical Chatbot with Local LLaMA 3 (Aug 2025): Engineered a fully offline, privacy-first chatbot using LLaMA 3, LangChain, and sentence-transformer embeddings to retrieve and answer queries from medical PDFs built a Flask backend and custom HTML/CSS/JS chat interface, achieving <1 sec response times while eliminating external API costs.

Advanced Machine Learning & Analytics

- **Ensemble & Unsupervised Learning Analysis (Nov 2024–Dec 2024):** Applied PCA & UMAP to reduce dimensions on Tortuga user-engagement and Wine Quality chemical datasets (silhouette = 0.99), then used GridSearchCV-tuned Random Forest & AdaBoost to uncover distinct behavior segments and key quality drivers.
- **Insurance Premium Estimator(MLOps)(June 2025–July 2025)** Web App Training a regression model ($R^2=0.75$) on health data deployed via a Streamlit UI with a FastAPI backend for real-time premium predictions

Data Visualization & Foundational Analytics

- **U.S. Police Shootings Analysis (Nov 2024–Dec 2024):** Created interactive Tableau dashboards to visualize nationwide trends, demographics, and geographic patterns with data storytelling.

EDUCATION

Hofstra University, Zarb School of Business

Jan 2024 – May 2025

M.S. in Business Analytics, GPA: 3.87/4.0

Hempstead, NY

Authorized to work in the U.S. under OPT & STEM OPT visa (expires May 2028)

Manipal University

Aug 2016 – May 2020

B.Tech in Computer Communication Engineering

Jaipur, India