HARSHIT JAIN

Bloomington, IN | jainharshitanil@gmail.com | +1 (930) 904-1020 | LinkedIn | GitHub

EDUCATION

Master of Science (Data Science) *August 2024 – May 2026*
Indiana University Bloomington CGPA: 3.73/4.0

Bachelor of Technology (Electronics & Telecommunication) *December 2020 – June 2024*
Dwarkadas J. Sanghvi College of Engineering, Mumbai CGPA: 8.69/10

- **Relevant Coursework:** Introduction to Statistics, Applied Machine Learning, Applied Algorithms, Data Mining, Applied Database Technologies, MGMT Access Use Big Data.

TECHNICAL SKILLS

- **Programming & Data Analysis:** Python (NumPy, Pandas, Scikit-learn), R, SQL
- **Machine Learning & AI:** Supervised & Unsupervised Learning, Deep Learning, NLP, LLMs (Transformers)
- **Statistical Modeling:** Regression, Time Series Analysis, Bayesian Methods
- **Cloud & Big Data Tools:** Google Cloud (BigQuery, Dataflow), AWS (S3, Lambda), Databricks
- **Data Workflow & Visualization:** Data Cleaning, Feature Engineering, Tableau, Power BI

WORK HISTORY

Machine Learning Intern, Feynn Labs *June 2023 – August 2023*

- Developed **machine learning models** to support **customer segmentation**, enabling more **personalized marketing strategies**.
- Built **Python-based data pipelines** for **data extraction**, **preprocessing**, and **transformation** from **structured data sources**.
- Conducted **A/B testing** on marketing campaigns and analyzed **user behavior metrics** to guide future campaign design.
- Automated **ETL workflows** using **Google Cloud Platform tools** (**BigQuery**, **Cloud Functions**), improving data pipeline **efficiency and scalability**.

Python Development Intern, Hackveda Limited *September 2022 – November 2022*

- Designed and automated **ETL pipelines** using **Python** and **SQL**, streamlining **data integration** across multiple sources.
- Built and deployed **AWS Lambda functions** to support **event-driven data processing** and reduce manual overhead.
- Implemented **validation checks** and **logging mechanisms** to enhance the **reliability** and **robustness** of internal data workflows.

PROJECTS

Cardiovascular Health Analysis Dashboard *March 2025 – May 2025*

- Built an end-to-end analytics pipeline on **Google Cloud Platform**, using **BigQuery** for SQL-based feature engineering and **Looker Studio** for interactive data visualization across 68,000+ health records.
- Identified key cardiovascular risk factors by applying **data wrangling**, **exploratory analysis**, and **dashboard storytelling**, enabling insights for both technical teams and non-technical stakeholders.

Neural Network-Based Image Analysis *August 2024 – October 2024*

- Analyzed **high-dimensional datasets** using **PCA, t-SNE, and LLE**, improving clustering accuracy by **25%**.
- Implemented **K-Means and Expectation Maximization algorithms**, refining data segmentation by **15%**.
- Conducted **Procrustes analysis** to align **machine-generated embeddings** with human judgments, enhancing model interpretability.