HARSHIT JAIN

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*
University of Mumbai, India   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

- **Machine Learning & AI**: Python (scikit-learn, TensorFlow, PyTorch), ML model development
- **Data Analysis & Processing**: SQL, Pandas, NumPy, Google BigQuery
- **Data Visualization**: Tableau, Power BI, Seaborn, Matplotlib
- **Cloud & Big Data**: AWS, Google Cloud, Snowflake
- **Product Analytics & A/B Testing**: Cohort analysis, statistical testing

WORK HISTORY

Research Volunteer, Data Science & AI Lab (Kelly School of Business, IU Bloomington)   *June 2025 – Present*
- Collaborating with PhD and MS researchers to integrate **large-scale cybersecurity datasets** (e.g., Hugging Face model metadata) into a **research-grade interactive dashboard** using **SQL** and **Python**.

Machine Learning Intern, Feynn Labs   *June 2023 – August 2023*
- **Developed SQL-based data pipelines** for data transformation and reporting, improving data processing efficiency by **30%**.
- **Designed Power BI dashboards** for customer analytics, enhancing **real-time decision-making** for business teams.
- Conducted **data validation and quality checks**, ensuring consistency across datasets.
- **Automated ETL workflows on Google Cloud (BigQuery, Cloud Functions)**, reducing data processing time by **35%**.

Python Development Intern, Hackveda Limited   *September 2022 – November 2022*
- Automated key processes, reducing manual errors by **25%** and improving efficiency.
- Developed scripts for **data extraction, validation, and SMTP email automation**, streamlining workflows.
- Deployed **AWS Lambda functions in Python**, optimizing database operations and image processing, cutting processing time by **30%**.

PROJECTS

Cardiovascular Health Analysis Dashboard   *Spring 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   *Fall 2024*
- Designed and trained **deep learning models for image feature extraction**, improving classification accuracy by 25%.
- Applied **PCA, t-SNE, and LLE** for high-dimensional image clustering.
- Developed a **convolutional neural network (CNN)** for feature recognition and segmentation.

Deep Learning-Based Air Pollution Prediction   *Summer 2023*
- Implemented a **deep learning model** for accurate air pollution forecasting using **historical air quality data**, **meteorological variables**, and **spatial-temporal dependencies**.
- Trained on a large dataset, the model effectively **predicts pollutant concentrations**, aiding in **pollution management** and **mitigation strategies**.