

Mahimna Darji

+1 (930)-333-4112 | darjimahimnads@gmail.com | [linkedin.com/in/mahimnadarji](https://www.linkedin.com/in/mahimnadarji) | github.com/MahimnaDarji

EDUCATION

Indiana University Bloomington

Bloomington, Indiana, USA

Master of Science in Data Science: **GPA: 3.73/4**

Aug 2023 – May 2025

- Courses: Applied Machine Learning, Data Mining, Data Visualization, Advanced Database Concepts, Usable Artificial Intelligence, Statistics, Social Media Mining, Software Engineering

Indus University

Ahmedabad, Gujarat, India

Bachelor of Technology in Computer Science: **GPA: 3.93/4**

Jul 2019 – Apr 2023

- Courses: Database Management System, Machine Learning, Data Structures & Algorithms, Object Oriented Programming, Operating Systems, Computer Networks, Soft Computing, Cloud Computing

SKILLS SUMMARY

Programming Languages : Python, R, JavaScript, C, C++

Database & Cloud Platforms : Google Cloud Platform, Amazon Web Services, Azure SQL, MySQL, PostgreSQL, MongoDB, BigQuery

Data Engineering & ETL : PySpark, Apache Spark, IBM DataStage, Pandas, Docker, Git, REST APIs, Alteryx

Data Visualization : Tableau, Power BI, Streamlit, Plotly, Matplotlib, Seaborn, Gephi, Excel, PowerPoint

Machine Learning & AI : TensorFlow, Neural Networks, LLMs, NLP, CNN, RAG, FAISS, LangChain, Hugging Face

EXPERIENCE

Data Scientist

Jan 2025 – Present

Project 990 Inc.

Lewis Center, Ohio, USA

- Build ETL pipelines** to collect, clean, and standardize financial grant data from 7 sources across 53 nonprofit organizations spanning 2013 to 2021, **reducing processing time by 40%** and enabling structured analysis using PySpark, SQL, and Python.
- Transform over 130,000 records for time-series and geospatial analysis** to identify grant distribution patterns and funding trends across regions and provide inputs for forecasting and resource allocation.
- Create Tableau dashboards to highlight donation trends and funding disparities. **Collaborate with executive leadership to define KPIs.** Document the data pipeline, regression testing logic, and version control through Git to ensure reproducibility.

Machine Learning Intern

May 2024 – Aug 2024

ManekTech

Ahmedabad, Gujarat, India

- Engineered a Python-based backend for real-time AQI forecasting using cloud-native infrastructure and **automated ingestion from REST APIs.** **Developed CI/CD workflows** using Git for streamlined deployment.
- Improved model accuracy by 7.6%** through hyperparameter tuning and feature selection using scikit-learn and NumPy. **Optimized preprocessing logic across 4 supervised learning models** including regression and decision trees, enabling more reliable predictions.
- Deployed the final model into a real-time pipeline** with custom monitoring scripts using Docker containerization. Enabled early alerts for hazardous conditions through data modeling techniques.

Data Science and Business Analytics Intern

Jul 2022–Sep 2022

Start Tech Academy

Ahmedabad, Gujarat, India

- Implemented an image classification model** using Convolutional Neural Networks and applied Natural Language Processing techniques to analyze 2,000 image-text pairs from educational materials. Enabled multimodal analysis and **improved prediction accuracy by 11%** through model optimization using TensorFlow frameworks and Transformers for text classification tasks.
- Built backend-to-frontend data pipelines** to track model performance and deliver insights. **Deployed Power BI dashboards** with interactive visualizations and user-friendly views to support stakeholder decision-making.

PROJECTS

Insightify - RAG Based AI Knowledge Assistant

June 2025 – Aug 2025

- Built an AI knowledge assistant** to transform unstructured enterprise documents into searchable insights using a full RAG pipeline. **Accomplished structured access to over 1000 PDFs, TXT, and emails** as measured by converting them into JSONL, applying smart chunking with LangChain, and generating embeddings with Sentence Transformers for semantically rich retrieval.
- Developed a FAISS vector index** with metadata for fast semantic search and integrated Hugging Face T5 models for grounded, context-aware answers. **Delivered 20% faster and more accurate** queries on weekly reports, emails, feedback, and transcripts as measured by contextual accuracy.

Supply Chain Optimization Platform

May 2025 – Jun 2025

- Designed an interactive supply chain analytics platform using Python and **simulated logistics across 50,000 shipment records** from a US-based warehouse network. Built demand forecasts using Prophet, performed delay detection using Z-score trends and shipment heatmaps using Folium.
- Created Stability Score and Efficiency Index to flag risky SKUs. Developed Tableau dashboards to visualize cost and delivery anomalies. Enabled decision-makers to compare vendor tradeoffs. **Improved on-time delivery predictions by 15%** through forecasting techniques and data modeling.

Los Angeles Police Department Crime Detection

Aug 2024 – Dec 2024

- Developed **scalable predictive models on 900,000 LAPD crime records** using stratified sampling, time series modeling, and spatial clustering in PyTorch and Spark. **Improved hotspot detection accuracy by 18%** and extracted key patterns from unstructured geospatial data.
- Visualized risk zones, victim demographics, and crime trends** across weeks and months using heatmaps and clustering techniques. Created a Streamlit dashboard to display visualizations, supporting location-based intelligence and behavioral trend analysis for strategic planning.

CERTIFICATIONS

Data Engineering (AWS), SQL Fundamentals (DataCamp), Data Science (BCG), Data Methodology (Coursera), Cloud Computing (AWS)