

Mahimna Darji

+1 (930)-333-4112 | darjimahimnads@gmail.com | linkedin.com/in/mahimnadari | github.com/MahimnaDarji

EDUCATION

Indiana University

Master of Science in Data Science: **GPA: 3.7/4**

Bloomington, Indiana, USA

Aug 2023 – May 2025

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

Indus University

Bachelor of Technology in Computer Science: **GPA: 3.9/4**

Ahmedabad, Gujarat, India

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 & Frameworks	: Python, R, Java, JavaScript, Scala, C++, React.js, Node.js, LangChain, LangGraph
Database & Cloud Platforms	: MySQL, PostgreSQL, Oracle PL/SQL, MongoDB, Azure SQL, AWS, Google Cloud Platform
Data Engineering & ETL	: PySpark, Apache Spark, Hadoop, IBM DataStage, Airflow, Docker, Kubernetes, Alteryx, REST APIs
Data Visualization	: Tableau, Power BI, Streamlit, Plotly, Matplotlib, Seaborn, Gephi
Machine Learning & AI	: TensorFlow, PyTorch, Hugging Face, LLMs, NLP, RAG, FAISS, Neural Networks, CNN, Feature Engineering

EXPERIENCE

Data Scientist

Project 990 Inc.

Jan 2025 – Present

Columbus, Ohio, USA

- Implement scalable ETL pipelines** using **PySpark, SQL** and **Python** to **ingest and structure** financial grant data from 7 different sources spanning from 2013 to 2021, **reducing processing time by 40%** and enabling **data warehousing**.
- Model and transform over 130,000 records**, executing **time-series** and **geospatial analysis** to identify grant distribution patterns for **resource allocation and forecasting**.
- Develop interactive Tableau dashboards** to visualize and communicate **donation trends and funding disparities**. **Consult** with executive leadership to **define KPIs** and ensure data reproducibility by **documenting** the pipeline, regression testing logic and **Git** version control.

Machine Learning Intern

May 2024 – Aug 2024

Ahmedabad, Gujarat, India

- ManekTech
- Engineered a Python based backend** for AQI forecasting with **automated data ingestion** from REST APIs and **developed** an interactive **React.js interface** for monitoring air quality trends.
 - Refined preprocessing logic** and **improved model accuracy by 7.6%** across 4 **supervised learning models** including **regression** and **decision trees**, through strategic **hyperparameter tuning** and **feature selection**.
 - Enabled** early alerts for hazardous conditions using **data modeling techniques** and **deployed** the trained model into a **production pipeline** with custom monitoring scripts using **Docker**.

Data Science and Business Analytics Intern

Jul 2022 – Sep 2022

Ahmedabad, Gujarat, India

- Start Tech Academy
- Built an image classification model** using **Convolutional Neural Networks** and **applied Natural Language Processing** techniques to **analyze 2,000 image-text pairs** from **educational materials**. **Optimized** the solution using **TensorFlow** and **Transformers**, **improving prediction accuracy by 11%** for text classification tasks.
 - Designed backend-to-frontend data pipelines** to track model performance and deliver insights. **Deployed Power BI dashboards** with interactive visualizations to support stakeholders in decision making.

PROJECTS

Insightify - RAG Based AI Knowledge Assistant | [GitHub](#)

- Designed** an **AI knowledge assistant** with a full **RAG pipeline** to convert **10,000+ unstructured enterprise documents** including PDFs, TXT and emails into searchable insights using **LangChain** for chunking and **Sentence Transformers** for embeddings.
- Created** a **FAISS vector index** with metadata for **fast semantic search**. **Integrated Hugging Face T5 models** for context-aware answers, delivering **20% faster, more accurate queries** on enterprise data.

Supply Chain Optimization Platform | [GitHub](#)

- Developed** an **interactive supply chain analytics platform** in **Python** and simulated logistics across **50,000 shipment records** from a US warehouse network. **Generated demand forecasts** using **Prophet**, **detected shipment delays** using **Z-score trends** and **created heatmaps** with **Folium**.
- Calculated Stability Score and Efficiency Index** to identify and monitor high-risk SKUs. **Designed Tableau dashboards** and **Excel pivot tables** for scenario-based analysis. **Increased on-time delivery prediction accuracy by 15%** through data forecasting and anomaly detection techniques.

Los Angeles Police Department Crime Detection | [GitHub](#)

- Built scalable predictive models** on **900,000 LAPD crime records** using **stratified sampling** and **spatial clustering**. **Applied data modeling techniques** to organize datasets, **improving hotspot detection accuracy by 18%** and identifying key patterns from **unstructured geospatial data**.
- Optimized distributed Apache Spark pipelines** with security measures and **applied heatmap** and **clustering methods** to study **risk zones, demographics and crime trends**. **Deployed results** through a **Streamlit dashboard** for location-based analysis and visualization.

CERTIFICATIONS

FinTech (IIT Delhi), Data Science Professional (Oracle), Data Engineer (AWS), Cloud Architect (AWS), Data Science (IBM), Data Analytics (BCG), SQL Fundamentals (DataCamp)