

Mahimna Darji

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

SUMMARY

Data Scientist with mid-level experience leading a three-person team and working directly with executive leadership, including the Lead Data Scientist and CEO, to share insights on donation trends and define KPIs. The analysis and dashboards I built at Project 990 Inc. contributed to the organization winning the ARNOVA President's Award. Alongside my work, I regularly write technical blogs on topics such as data science, data engineering, LLMs, and storage systems. I focus on presenting technical analysis in a way that is easy to understand for both technical and non-technical audiences.

EDUCATION

Indiana University

Bloomington, Indiana, USA

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

Aug 2023 – May 2025

Indus University

Ahmedabad, Gujarat, India

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

Jul 2019 – Apr 2023

EXPERIENCE

Data Scientist

Jan 2025 – Present

Project 990 Inc.

Columbus, Ohio, USA

- **Implement scalable data pipelines to ingest financial grant data** from 7 different sources spanning from 2013 to 2021, **reducing processing time by 40%** and enabling data warehousing.
- **Use SQL to merge and validate datasets**, then apply Python to **model and transform over 130,000 records**, performing **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**. Ensure data reproducibility by **documenting data pipelines** and maintaining **version control using Git**.

Machine Learning Intern

May 2024 – Aug 2024

ManekTech

Ahmedabad, Gujarat, India

- **Built a Python backend** for AQI forecasting with **automated data ingestion** from REST APIs and developed a React.js interface to monitor air quality trends in real time.
- **Refined preprocessing logic** for Linear Regression and Decision Tree models, **then improved model accuracy by 7.6%** through model optimization methods such as **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

Start Tech Academy

Ahmedabad, Gujarat, India

- **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 used for 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.

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.

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, Salesforce
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

CERTIFICATIONS

Data Science Professional (**Oracle**), Data Science (**IBM**), Data Engineer (**AWS**), Data Analytics (**BCG**), SQL Fundamentals (**DataCamp**)