Mohammed Pathariya

in linkedin.com/in/mjpathariya | 🌐 mjpathariya.com | github.com/MohammedPathariya
✉ mjpathariya7@gmail.com | 📞 +1 (930) 333-7212 | 📍 Indiana, USA

EDUCATION

Master of Science in Data Science, Indiana University — Bloomington, IN, USA
GPA: 3.6/4.0 — Coursework: LLMs, Deep Learning, MLOps, Cloud Computing — *Aug. 2024 – Exp. May. 2026*

BE in Artificial Intelligence & Data Science, Pune University — Pune, MH, India
GPA: 3.8/4.0 — ML, Data Structures & Algorithms, Operating Systems, Statistics — *Aug. 2020 – May. 2024*

EXPERIENCE

Graduate Research Assistant, AI in Sports Analytics — Aug. 2025 – Present
Indiana University, Kelley School of Business — *Bloomington, IN*

- Engineered an end-to-end data pipeline in **Python** to process YouTube content, automating audio extraction, transcription, and speaker diarization (**WhisperX**) and reducing manual data collection effort by **95%**.
- Architected a semantic reconstruction method using **Sentence-BERT** embeddings and an NLP feature layer with **VADER** to transform unstructured, diarized text into an analysis-ready dataset for predictive modeling.
- Spearheaded a comprehensive data integrity audit on the NBA Playoff dataset, identifying and documenting critical data gaps to ensure source data reliability and completeness.

Data Engineering Intern — Feb. 2022 – Jul. 2022
Sparkwood IT Solutions — *Pune, India*

- Orchestrated daily ETL pipelines using **Python**, **SQL**, and **Airflow**, building **pandas** jobs to transform 10GB+ of weekly data for a centralized **PostgreSQL** data warehouse.
- Reduced query times by **40%** by optimizing SQL schemas with strategic indexing, directly increasing the productivity and efficiency of downstream data consumers.
- Partnered with data and engineering teams to define data product requirements that aligned with cross-functional needs for reporting and feature development.

PROJECTS

The Digital Forge – AI Agents for Code Automation | *Multi-Agent Systems, LLMs* — May. 2025 – Aug. 2025

- Developed an autonomous multi-agent system that cuts boilerplate coding time by **60%** with an **85% success rate** for Python script generation.
- Engineered the agent collaboration framework (prompt chaining, stateful memory) and automated validation using a **Docker** container, **GitHub Actions** CI/CD, and a **Streamlit** UI.
- Implemented a self-correction loop that used test feedback to autonomously debug code, successfully correcting over **70%** of initial runtime errors.

LearnLoop – AI Study Companion | *LLMs, Full-Stack, MLOps, EdTech* — Jan. 2025 – Apr. 2025

- Improved user knowledge retention by **25%** with a scalable GPT-powered learning assistant architected to handle **500+ concurrent users**.
- Automated deployment with a **GitHub Actions CI/CD pipeline** and analyzed user session logs to monitor API performance and inform the product roadmap.

AudioGroove – AI Music Composition Platform | *Deep Learning, Data Pipelines* — Oct. 2023 – Apr. 2024

- Achieved a cross-entropy loss of **0.78** with a **PyTorch** LSTM model on a generative AI platform that creates royalty-free music.
- Managed the end-to-end MLOps lifecycle: from processing 17K+ MIDI files with a **Dask** pipeline and tracking **50+ runs** in **MLflow** to deploying the final containerized **Flask** API.

TECHNICAL SKILLS

Languages: Python, SQL, R, Bash
Data Science & Statistics: Pandas, NumPy, Dask, A/B Testing, Statistical Inference, Statsmodels
Machine Learning & Deep Learning: scikit-learn, PyTorch, Hugging Face Transformers
AI Development & MLOps: LangChain, CrewAI, WhisperX, MLflow
Cloud & Data Engineering: GCP (Vertex AI, Cloud Storage), PostgreSQL, Apache Airflow
MLOps & Deployment: Docker, Flask, React, Streamlit, REST APIs, Hugging Face, Vercel
Developer Tools: Git, GitHub, GitHub Actions, VS Code, Jupyter Notebooks

PUBLICATIONS

Tunes by Technology: Comprehensive Survey of Music Generation Models — Apr. 2024
IEEE ICC Robins — Published a comparative analysis of GANs/RNNs for music composition. — *Coimbatore, TN, India*