

Mohammed Pathariya

🌐 [linkedin.com/in/mjpathariya](https://www.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 <i>GPA: 3.6/4.0 — Coursework: LLMs, Deep Learning, MLOps, Cloud Computing</i>	Bloomington, IN, USA <i>Aug. 2024 – Exp. May. 2026</i>
BE in Artificial Intelligence & Data Science, Pune University <i>GPA: 3.8/4.0 — ML, Data Structures & Algorithms, Operating Systems, Statistics</i>	Pune, MH, India <i>Aug. 2020 – May. 2024</i>

TECHNICAL SKILLS

Languages: Python, SQL, R, Bash
Data Science & Statistics: Pandas, NumPy, A/B Testing, Statistical Inference, Statsmodels
Machine Learning & MLOps: scikit-learn, PyTorch, Dask, MLflow, LangChain, CrewAI
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

PROJECTS

- The Digital Forge – AI Agents for Code Automation** | *Multi-Agent Systems, LLMs* May. 2025 – Aug. 2025
- Reduced boilerplate coding time by **60%** by developing an autonomous multi-agent system that achieved an **85% success rate** in generating functional Python scripts from prompts.
 - Engineered robust agent collaboration using prompt chaining, stateful memory, and tool augmentation for file system and shell access.
 - Containerized the agent environment with **Docker** and used **GitHub Actions** for automated testing, ensuring reproducibility and enabling stakeholder validation via a Streamlit UI.
- LearnLoop – AI Study Companion** | *LLMs, Full-Stack, MLOps, EdTech* Jan. 2025 – Apr. 2025
- Improved user knowledge retention by **25%** by architecting a scalable, GPT-powered learning assistant designed to handle **500+ concurrent users**.
 - Automated the full build-and-deploy process using a **CI/CD pipeline in GitHub Actions** and implemented API logging and performance monitoring to track latency and error rates.
 - Analyzed user session logs to generate insights that directly informed the product roadmap and prioritized future content and feature development.
- AudioGroove – AI Music Composition Platform** | *Deep Learning, Data Pipelines* Oct. 2023 – Apr. 2024
- Empowered content creators with royalty-free music by engineering a generative AI platform using a PyTorch LSTM model that achieved a cross-entropy loss of **0.78**.
 - Managed the full model lifecycle, from building a parallelized **Dask** data pipeline (17K+ MIDI files) to tracking **50+ training runs** with **MLflow**.
 - Deployed a decoupled, containerized Flask backend on Hugging Face Spaces, providing a responsive and cost-efficient user-facing service.

EXPERIENCE

- Data Engineering Intern** Feb. 2022 – Jul. 2022
Sparkwood IT Solutions Pune, MH, India
- Orchestrated daily ETL pipelines using **Python, SQL, and Airflow** to ingest data from Salesforce and application logs into a centralized **PostgreSQL** data warehouse.
 - Reduced query times by **40%** by optimizing SQL schemas with strategic indexing, directly increasing the productivity of downstream data consumers.
 - Built pandas transformation jobs to reliably process over 10 GB of data weekly, ensuring data quality for critical analytics and modeling.
 - Collaborated with data and engineering teams to align data products with cross-functional requirements for reporting and feature development.

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