Kratik Rathi

+1 (812) 822-7764 | rathi.kratik08@gmail.com | linkedin.com/kratikrathi | Kratik-Rathi-Portfolio | github.com/Kratik-Rathi

EDUCATION

Indiana University - Bloomington **August 2024 – May 2026**
Master of Science in Data Science GPA: 3.6
Coursework: Information Visualization, Applied Database Technologies, Applied Machine Learning, Intro to Statistics, Management Access of Big Data

Medi-Caps University **August 2019 – July 2023**
Bachelor of Technology in Computer Science and Engineering GPA: 3.4
Coursework: Data Structures and Algorithms, Big Data Engineering, Data Mining, Data Visualization, Software Engineering

TECHNICAL SKILLS

Programming Languages: Python, R, SQL, C++, C#
Databases & Warehousing: Snowflake, BigQuery, MySQL, PostgreSQL, MS SQL Server, MongoDB
Cloud Platforms & Big Data Technologies: Databricks, Apache Spark (PySpark), AWS, GCP
BI & Data Visualization: Power BI (DAX, Power Query), Tableau, Looker Studio, Excel
Libraries & Frameworks: Pandas, NumPy, Scikit-learn, PyTorch, TensorFlow, Keras, OpenCV, LangChain, Matplotlib, Seaborn, Streamlit, Flask, .NET
Machine Learning: Regression Models, Tree-based Models (XGBoost, Random Forest), SVM, Clustering (K-Means, GMM), Deep Learning (CNN)
Development Tools: Git (GitHub, GitLab), Docker, Postman

WORK EXPERIENCE

Associate Instructor at **Indiana University |** *Bloomington, IN* **May 2025 – Present**
- Delivered NLP and Database lectures to 40+ graduate students on transformer models, LLMs, prompt engineering and advanced database concepts.
- Led collaborative sessions that clarified complex concepts, earning a 90% recommendation rate based on student response distribution.

Software Engineer Intern at **Genmark AI Inc. |** *Remote, USA* **July 2025 – August 2025**
- Developed the initial version of Gemini-powered RAG pipeline using Python and LangChain, embedding crawled GCP data into a Milvus vector database and containerizing the end-to-end system with Docker to support scalable, context-aware retrieval.
- Enabled approximately 60% faster content generation for clients by implementing a chat-based interface over embedded assets, producing accurate, context-aware responses using Gemini and a Dockerized vector store.
- Reduced content creation time to under 5 minutes by building a tool that analyzed top-20 trending TikTok videos, generated transcripts via Whisper, and produced personalized videos using Google Veo2; recognized as “Most Innovative Project”.

Research Assistant at **Indiana University – Indiana Innocence Project |** *Bloomington, IN* **January 2025 – May 2025**
- Constructed a Python-based ingestion pipeline using BeautifulSoup to scrape and normalize legal case data from intake forms and judicial sources into a PostgreSQL database, cutting manual data entry time from 30 minutes to 1 minute.
- Designed a Looker Studio dashboard to visualize case timelines, severity and procedural status, helping assign review priority to 200+ cases.

Data Analyst Intern at **Swastika Investmart Ltd. |** *Indore, India* **January 2024 – July 2024**
- Created an automated analytics workflow using Python, Flask API, SQL and cron jobs to generate daily reports on 50K+ records, providing product-level and campaign-level insights across client zones, which led to a 20% increase in conversions and reduced manual effort by 90%.
- Engineered a scalable Docker-based scheduler to centralize deployment of internal APIs and integrated Prometheus with unified logging, reducing debugging time by 60% and streamlining future rollouts.

Data Analyst Intern at **Mahindra and Mahindra Ltd. |** *Mumbai, India* **January 2023 – April 2023**
- Implemented a time-series forecasting model using ARIMAX and SARIMAX to predict financial trends for the next 6 years with 87% accuracy using past 20 years of monthly microeconomic data, which helped forecast sales volume across multiple car models.
- Compiled and processed a 30K+ record dataset across 20+ car models to analyze part cross-utilization, supporting cost optimization efforts.

Data Analyst Intern at **Tech Mahindra Ltd. |** *Pune, India* **June 2022 – August 2022**
- Cleaned and transformed CRM data to improve usability and built interactive Tableau dashboards to deliver real-time visibility of customer deliverables, enabling improved status tracking and reporting accuracy.

PROJECTS

IndyGo – Ridership Dashboard (Client Project, ENGR-E 583: Information Visualization) **May 2025**
- Ensured ridership accuracy by validating 38K GPS and APC records, eliminating 3% invalid entries through trip and route-specific thresholds.
- Designed an interactive Power BI dashboard analyzing 35K+ valid records, surfacing weekday vs weekend travel patterns, stop-level passenger flows, and route-level ridership KPIs, which led to a 3% improvement in service planning KPIs.

DocuVerse – ChatBot **January 2025**
- Developed a RAG-based document processing app with LangChain and Hugging Face embeddings to automate text chunking and semantic representation, reducing manual document review time by 80%.
- Enhanced retrieval with FAISS and integrated Gemma2-9b-it to enable context-aware Q&A and structured summarization. Implemented persistent chat history and deployed the system on Streamlit for real-time usage.

Lung X-ray Images Classification **December 2024**
- Performed dimensionality reduction on 1,227 X-ray images using PCA (retaining 90% variance) and applied K-Means and Gaussian Mixture clustering to visualize clusters in 2D for pattern separation.
- Achieved 98% classification accuracy on COVID-19, Pneumonia, and Normal cases by training and optimizing a CNN model with TensorFlow and Keras, supporting automated diagnosis from chest X-rays.

CERTIFICATIONS

AWS Certified Cloud Practitioner (January 2026)