

Kratik Rath

+1 (812) 822-7764 | krarathi@iu.edu | [linkedin.com/kratikrathi](https://www.linkedin.com/kratikrathi) | github.com/Kratik-Rathi | [Kratik-Rathi-Portfolio](#)

EDUCATION

Indiana University - Bloomington

Master of Science in Data Science

August 2024 – May 2026

Bloomington, Indiana

Medi-Caps University

Bachelor of Technology in Computer Science and Engineering

August 2019 – July 2023

Indore, India

TECHNICAL SKILLS

Languages: Python, C++, R, C#

Databases: SQL, MySQL, PostgreSQL, Microsoft SQL Server, SQLite, MongoDB

Libraries and Frameworks: TensorFlow, Keras, OpenCV, PyTorch, LangChain, Matplotlib, Pandas, Numpy, Pyodbc, Openpyxl, Scikit-learn, Seaborn, Streamlit, Flask Framework, .NET Framework

Machine Learning: CNN, Linear Regression, Logistic Regression, Decision Trees, Random Forest, XGBoost, SVM, Naive Bayes, K-Means, DBSCAN, Gaussian Mixture, Arimax, Sarimax

Tools/Technologies: Tableau, Power BI, Looker Studio, Microsoft Office, GitLab, GitHub, Git, Docker, Postman, AWS

Natural Language Processing (NLP): Hugging Face, GroqCloud, Retrieval-Augmented Generation (RAG)

WORK EXPERIENCE

Research Assistant

Indiana University - Indiana Innocence Project

January 2025 – Present

Bloomington, Indiana

- Developed a fully automated pipeline using BeautifulSoup to extract and upload a convict's case data from intake forms and publicly available judicial case details into a database, minimizing manual work by 75% for the team.
- Designed dashboard using Looker Studio to visualize case timelines, severity, and procedural status, enabling the team to efficiently sort and assess 200+ conviction cases.

Data Analyst Intern

Swastika Investmart Ltd.

January 2024 – July 2024

Indore, India

- Automated digital marketing performance reports using Flask API and cronjobs, daily analyzing 5M+ records across client zones to track product-level and campaign-level conversions over the past 90 days, thus improving organization decision-making and reducing manual effort by 90%.
- Programmed an ASP.NET API to automate daily checks for pending client status in the MutualFunds database over the past 7 days and triggered automatic WhatsApp reminders, improving process efficiency by almost 70%.
- Deployed all internal APIs on a remote server by creating a docker-based scheduler with scalability for future deployments. Improved monitoring with Prometheus and reduced debugging time by 60% through unified logging.

Data Analyst Intern

Mahindra and Mahindra Ltd.

January 2023 – April 2023

Mumbai, India

- Constructed a time-series forecasting model using 20 years of monthly microeconomic data to predict financial trends for the next 6 years with 87% accuracy.
- Created a dataset of 30,000+ records covering 20+ car models to analyze part usage to support a cost reduction project.

Student Trainee - Analytics

Tech Mahindra Ltd.

June 2022 – August 2022

Pune, India

- Cleaned and transformed CRM data to improve usability and developed interactive dashboards using Tableau to support customer engagement strategy planning.

PROJECTS

DocVerse - ChatBot

January 2025

- Made a RAG-based document processing app using LangChain for text chunking and Hugging Face embeddings, reducing processing time by 30% and manual analysis time by 80%.
- Implemented FAISS for fast similarity search and integrated Gemma2-9b-it for context-aware responses, enabling structured summarization and persistent chat history. Deployed on Streamlit for real-time query handling.

Lung Xray Images Classification

December 2024

- Performed analysis and clustering on 1,227 X-ray images using PCA for dimensionality reduction and with K-Means and Gaussian Mixture for visualizing separately in 2D while retaining 90% variance.
- Applied a CNN model to classify COVID-19, Pneumonia, and Normal cases, attaining 98% accuracy by leveraging TensorFlow and Keras for training and optimization.

Time-series forecasting on Air Passengers data

April 2023

- Built time-series forecasting models using ARIMAX and SARIMAX to predict monthly air passenger traffic with 97% accuracy; leveraged ACF and PACF analysis and automated model selection using AutoARIMA.