

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

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, Microsoft Office, GitLab, GitHub, Git, Docker, Postman

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

WORK EXPERIENCE

Research Assistant

January 2025 – Present

Indiana University - Department of Criminology and Criminal Justice

Bloomington, Indiana

- Developing a database for over **2,000** historical wrongful conviction cases, applying **data mining** techniques to digitize and analyze past requests, thus improving case selection efficiency and automating categorization processes.
- Implementing **deep learning** to automate case reviews and using **NLP** to analyze letters from applicants to enhance the accuracy of eligibility determination.

Data Analyst Intern

January 2024 – July 2024

Swastika Investmart Ltd.

Indore, India

- Automated digital marketing reports using **Flask API** and **Python** analyzing **5 million records** daily across all client-zones and scheduled these analytical reports via email using **cronjobs** for timely stakeholders updates.
- Developed an **ASP.NET API** to verify pending client status in the MutualFunds database and update records on Netcore, eliminating manual effort by 100%. The API also generated push reference logs for accurate data verification.
- Designed a **Docker-based** universal scheduler to centralize API monitoring and logging, hosted **25 APIs** with the capacity to scale for additional API deployments, ensuring efficient data pipeline orchestration.

Data Analyst Intern

January 2023 – April 2023

Mahindra and Mahindra Ltd.

Mumbai, India

- Constructed **ARIMAX** and **SARIMAX** time-series forecasting models using **Python** on 20 years of monthly microeconomic regression data (2001-2020), achieving **87% accuracy** in predicting financial trends for the next 6 years.
- Created a matrix of **3000+ data** on **MS Excel** to analyze part usage, continuity, and discontinuity across car models.

Student Trainee - Analytics

June 2022 – August 2022

Tech Mahindra Ltd.

Pune, India

- Examined and transformed CRM data using **Microsoft Excel** to improve data accuracy and relevance. Produced **5 interactive dashboards** in **Tableau** to enhance data visualization and insights, which facilitated strategic planning.

PROJECTS

DocVerse - ChatBot

January 2025

- Built 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, with **K-Means** and **Gaussian Mixture** for visualizing separately in 2D while retaining **90%** variance.
- Applied a **deep learning model** to classify COVID-19, Pneumonia, and Normal cases, attaining **98%-99% accuracy** by leveraging **TensorFlow/Keras** for training and optimization.

Time-series forecasting on Air Passengers data

April 2023

- Programmed a predictive model for monthly air passenger traffic with **97%-98% accuracy** using **ARIMAX** and **SARIMAX**. Analyzed seasonality and continuity through **ACF/PACF**, and optimized model selection with **AutoARIMA** for more accurate forecasting.