Kratik Rathi

+1 (812) 822-7764 | krarathi@iu.edu | linkedin.com/kratikrathi | github.com/Kratik-Rathi | Kratik-Rathi-Portfolio

EDUCATION

Indiana University - Bloomington   **August 2024 – May 2026**
Master of Science in Data Science - GPA 3.5   *Bloomington, Indiana*

Medi-Caps University   **August 2019 – July 2023**
Bachelor of Technology in Computer Science and Engineering - GPA 3.43   *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   **January 2025 – May 2025**
Indiana University - Indiana Innocence Project   *Bloomington, Indiana*

- Developed a pipeline to automate extraction and integration of case data from intake forms and judicial sources into a database using BeautifulSoup, which mitigated manual data entry time by 75% and accelerated case review workflows.
- Designed a Looker Studio dashboard to visualize case timelines, severity, and procedural status, helping the team to decide the review priority of 200+ cases.

Data Analyst Intern   **January 2024 – July 2024**
Swastika Investmart Ltd.   *Indore, India*

- Automated digital marketing performance reports using Flask API and cronjobs, daily analyzing 50K+ records across client zones to track product-level and campaign-level conversions over the past 90 days, resulting in a 20% increase in conversions per zone and reducing manual effort by 90%.
- Programmed an ASP.NET API to identify clients with pending mandate statuses from over the past 7 days and trigger WhatsApp reminders, boosted on-boarding speed and improved process efficiency by 70%.
- Engineered a scalable Docker-based scheduler to deploy all previously scattered internal APIs on a remote server, enabling seamless future deployments. Integrated Prometheus and unified logging to cut debugging time by 60%.

Data Analyst Intern   **January 2023 – April 2023**
Mahindra and Mahindra Ltd.   *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, which helped forecast sales volume across multiple car models.
- Created a 30K+ records dataset covering 20+ car models to analyze cross utilization of car parts to support cost optimization.

Student Trainee - Analytics   **June 2022 – August 2022**
Tech Mahindra Ltd.   *Pune, India*

- Cleaned and transformed CRM data to improve usability and built interactive Tableau dashboards to deliver real-time visibility of customer deliverables through status tracking.

PROJECTS

DocVerse - ChatBot   **January 2025**

- Made a RAG-based document processing app using LangChain for text chunking and Hugging Face embeddings, reducing processing 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.