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 *Bloomington, Indiana*

Medi-Caps University **August 2019 – July 2023**
Bachelor of Technology in Computer Science and Engineering *Indore, India*

TECHNICAL SKILLS

Programming Languages: Python, C++, R, C#
Databases: SQL, MySQL, PostgreSQL, Microsoft SQL Server, SQLite
Libraries and Frameworks: TensorFlow, Keras, CNN, OpenCV, PyTorch, LangChain, Matplotlib, Pandas, Numpy, Pyodbc, Openpyxl, Scikit-learn, Seaborn, Arimax and Sarimax, Streamlit, Flask API, ASP.NET
Machine Learning Libraries: Deep Learning Techniques (CNN), Supervised Learning Algorithms (Linear Regression, Logistic Regression, Decision Trees, Random Forest, Boosting, SVM, and Naive Bayes), Clustering Techniques(K-Means, DBSCAN, Gaussian Mixture), Time-series Forecasting Techniques(Arimax, Sarimax)
Tools / Technologies: Tableau, Power BI, Microsoft Office, GitLab, GitHub, Git, Docker, Postman, Hugging Face, GroqCloud

WORK EXPERIENCE

Indiana University - Department of Criminology and Criminal Justice **January 2025 – Present**
Research Assistant *Bloomington, Indiana*

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

Swastika Investmart Ltd. **January 2024 – July 2024**
Software Engineer Intern *Indore, India*

- Automated Digital Marketing reports using **Flask** API and **Python** which analyzed **5 million records** daily of potential clients from different zones and scheduled reports on mail at midnight via **cronjob**.
- Developed an **ASP.NET** API to check the pending status of clients in the MutualFunds database for the past 7 days, update client records on Netcore, and generate logs with push reference numbers for accurate tracking and verification.
- Designed a **Docker-based** universal scheduler to manage and log 25 APIs on a single server, streamlining operations and enabling future API scheduling and log management through a single, scalable system.

Mahindra and Mahindra Ltd. **January 2023 – April 2023**
Data Analytics Intern *Mumbai, India*

- Constructed **ARIMAX** and **SARIMAX** time-series forecasting models using **Python** on 20 years of monthly regression data (2001-2020), achieving **87% accuracy** in predicting financial trends for the next 6 years.
- Created matrix of **1k+ data** on **MS Excel** for different car models to enhance individual tracking of parts.

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

- Acquired in-depth knowledge of CRM, E2E Billing, OSS, and BSS systems.
- Examined CRM data using MS Excel and enhanced data interpretability by creating **5 interactive dashboards** in **Tableau** enhancing insights for better decision-making.

PROJECTS

DocVerse - ChatBot **January 2025**

- Built a RAG-based application for PDF/Word file processing, extracting text into vectorized chunks using **LangChain** and **Hugging Face embeddings**, reducing document processing time by **30%** and manual analysis time by **80%**.
- Made an efficient pipeline to split and embed large texts, and implemented **FAISS** for similarity search. Integrated **Llama3-8b-8192 model** to generate accurate, context-aware responses, storing summaries and chat history for complete document processing and 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.
- Developed a **deep learning model** for classifying 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**

- Forecasted monthly air passenger traffic (1949-1960) using **ARIMAX** and **SARIMAX** with **97%-98% accuracy**.
- Checked seasonality and continuity using data **ACF/PACF** and tested models with **AutoARIMA** for the best fit.