Aaryan Manish Purohit

aaryan.prof1@gmail.com | linkedin.com/AaryanPurohit | 9309043794 | github.com/AaryanPurohit

PROFESSIONAL SUMMARY

Data Scientist with expertise in Python, ML, DL, SQL, BI tools, and Generative AI, developed through internships at Great Lakes IoM, IIT Patna, and Advertising Saga. Published two IEEE papers on preprocessing module (GARP) and OCR automation (CRBC). Developed 10+ projects, including a brain tumor classifier, an LLM-powered financial assistant, and a restaurant analytics solution that tripled sales. Focused on using machine learning and data analysis to solve real-world problems in healthcare, business and more.

EDUCATION

Indiana University Bloomington — Indiana, USA
Master of Science in Data Science - GPA: 4.0 — Aug 2024 - May 2026

Sardar Patel Institute of Technology — Mumbai, India
Bachelor of Technology in Computer Engineering - GPA: 3.8 — Feb 2020 - May 2024

TECHNICAL SKILLS

Programming & Visualization: Python, SQL, R, Power BI, Tableau, Looker Studio, Matplotlib, Seaborn, Plotly
Data Engineering & Cloud: AWS (EC2, S3, Lambda), Snowflake, MySQL, PostgreSQL, MongoDB, Docker, Kubernetes, Kafka
Generative AI & NLP: LangChain, OpenAI API, GPT-2, Llama, ChromaDB, Hugging Face Transformers, SpaCy, NLTK
ML/DL Algorithms & MLOps: PyTorch, TensorFlow, Keras, OpenCV, Regression, Random Forest, CNN, ResNet, MLflow, Git

EXPERIENCE

Great Lakes Institute of Management | ***Deep Learning Research Intern*** — Jun 2023 - Jul 2023

- Led a team of 4 members to design an algorithm for extracting relevant frames from live video feeds, improving efficiency by **20%** using Savgol filter and difference thresholds in Python, NumPy, and OpenCV for faster and accurate predictions.
- Implemented GPT and BERT models in a chatbot to analyze 100,000+ customer interactions, driving a **20%** increase in customer satisfaction and a **40%** reduction in response times.

Indian Institute of Technology (IIT Patna) | ***Deep Learning Research Intern*** — Jan 2023 - Jun 2023

- Researched and formulated a deep-learning model for brain tumor classification, harnessing a dataset of 7,000+ Brain MRI images with TensorFlow and Python, resulting in a model accuracy improvement of **1.5%** over previous benchmarks.
- Applied advanced models including VGG-19, Transformers, ResNet-50, and MobileNet with custom layers, alongside image enhancement techniques like object-centric extraction, gamma correction, and CLAHE, achieving top accuracy of **98.5%**.
- Partnered with IGIMS State Cancer Hospital, incorporating feedback from 4 radiologists and utilizing their patient dataset, which led to a research submission to Elsevier journal.

Advertising Saga | ***Data Analyst Intern*** — Jan 2022 - May 2022

- Automated ETL workflows using AWS Lambda and DynamoDB to streamline Amazon Seller data analysis, reducing manual handling time by **30%**, and used Power BI to enhance ad productivity analysis.
- Orchestrated Amazon Seller API integration to extract and visualize metrics, boosting ad campaign efficiency by **15%**.

PROJECTS AND PUBLICATIONS

FinEdu LLM: Financial Education Q&A Assistant — Nov 2024
Tech: Python, JSON, Transformers, LangChain, SentenceTransformers, LlamaIndex, Gradio

- Created a JSON dataset with financial terms, definitions, usage, and implications, and formulated FinEdu by leveraging the Llama-2-7b model and sentence-transformers embeddings for advanced semantic search & natural language understanding.
- Employed vectorized indexing with LangChain and LlamaIndex for scalable storage and retrieval of financial terminology, and deployed a Gradio-powered web interface with prompt engineering for real-time, user-friendly access to insights.

Zomato (Food Delivery Platform) Data Analysis for Restaurant — May 2024
Tech: Python, R, Power BI, SQL, HTML, CSS

- Analyzed 8 months of real-world restaurant Zomato data with 10,000+ records, identifying KPIs like top-selling categories, high-rated items, top 3 items per high-sales category, and forecasting seasonal trends with the help of Python, SQL & Power BI.
- Designed Sankey diagrams to map customer journeys and visualize low-rating orders, revealing reasons for negative feedback (e.g., missing items, poor quality, low quantity) using data-driven insights, displayed through an interactive HTML dashboard.
- Spearheaded actionable changes based on data analysis, by collaborating with cross-functional teams, including kitchen staff and delivery partners, which improved ratings from **3.6 to 4.1** and led to a **3x boost in sales**.

Crop and Weed Segmentation via ResNet-UNET Architecture — Jan 2024
Tech: TensorFlow, TensorFlow Lite, PyTorch, Python, OpenCV

- Developed a ResNet-34 U-Net model for UAV-based soil and weed segmentation, attaining **92.9%** Dice Coefficient, with research published in **IEEE Proceedings** and effectively presented through written and verbal communication at conferences.
- Constructed a dual-stage crop analysis framework, obtaining **99%** and **97.8%** accuracy for early and advanced growth stages respectively, facilitating precise weed distribution mapping and optimized crop management strategies.