Aaryan Manish Purohit

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

PROFESSIONAL SUMMARY

Data Analyst with experience in Python, R, SQL, and Power BI, developed through internships at Great Lakes Institute of Management, IIT Patna, and Advertising Saga. Published two IEEE papers on preprocessing module (GARP) and OCR automation (CRBC). Developed 10+ projects, including a T20 World Cup Cricket Team Analyzer and a Restaurant Analytics solution that tripled sales. Focused on using machine learning and data analytics to solve real-world problems across 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 & Analysis: Python, SQL, R, Advanced Excel, Google Sheets, DAX, Power Query, HTML, CSS
Data Visualization: Power BI, Tableau, Looker Studio, Matplotlib, Seaborn, Plotly
Databases & Cloud: MySQL, PostgreSQL, Snowflake, MongoDB, AWS (DynamoDB, S3, Lambda), Docker, Kafka
ETL & Data Wrangling: Pandas, NumPy, Power Query, BeautifulSoup, Scrapy
Machine Learning: PyTorch, TensorFlow, Keras, OpenCV, Regression, Random Forest, CNN, ResNet, MLflow, GitHub
Microsoft Suite: Word, Excel, PowerBI, PowerPoint, Outlook, Teams

EXPERIENCE

Great Lakes Institute of Management | *ML/DL 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.
- Mastered clustering methodologies to analyze a hotel customer dataset with 83,590 rows and 31 columns, contributing with cross-functional teams to identify key customer segments and patterns, which improved targeted marketing by **25%**.

Indian Institute of Technology (IIT Patna) | *ML/DL 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

Cricket Team Analyzer: T20 World Cup 2022   Jan 2025
Tech: Power BI, Python, JSON, JavaScript, Excel, DAX, Power Query, BeautifulSoup, Scrapy, Pandas
- Developed a Power BI dashboard with Power Query and DAX to analyze 200+ players, cutting team selection time by **40%** and improving decision-making with real-time projections and a drag-and-drop team builder.
- Scraped and processed 10,000+ data points from ESPN Cricinfo using Python (BeautifulSoup, Scrapy) and SQL, planning the best World Cup squad with accurate player evaluations, ensuring a team capable of 180+ run average and defending 150 runs.

Zomato (Food Delivery Platform) Data Analysis for Restaurant   May 2024
Tech: Python, R, Excel, Power BI, SQL, HTML, CSS
- Analyzed 8 months of real-world restaurant Zomato data with 10,000+ records, identifying KPIs like top-selling categories, 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**.

GARP - A Hybrid Preprocessing Module for Semantic Segmentation   Aug 2023
Tech: Python, PyTorch, TensorFlow
- Introduced a hybrid preprocessing module, GARP integrating Gamma Correction, Laplacian Kernel, and Histogram Equalization, achieving a **63.2%** reduction in AER and an 8.6% improvement in mIoU compared to unprocessed images.
- Evaluated GARP with Jaccard, Focal, and Binary Cross-Entropy loss functions to show superior performance over existing techniques. Presented and published the work in the IEEE Conference on Futuristic Technologies.