UTKARSH MATHUR

(425) 633-5220 • *utkarsh.mathur@gmail.com* • *LinkedIn* • *GitHub* • *Portfolio Website* • Greater Seattle Area, WA

Data Science and Software Engineering professional with 2+ years of experience in developing end-to-end solutions with Front-end and Back-end systems, Machine Learning & Deep Learning models, LLMs, Data Pipelines (ETL/ELT), and Data Analysis.

EDUCATION

University at Buffalo, State University of New York (UB) | New York, USA   January 2023 – June 2024
Master of Science in Data Science | *Capstone Project*: "*Graph Neural Networks and Large Language Models: A Literature Review*"

Indian Institute of Technology Roorkee (IIT Roorkee) | Uttarakhand, India   July 2018 – July 2022
Bachelor of Technology in Polymer Science (Chemical Engineering) | *Capstone Project*: "*Production of Sustainable Aviation Fuel*"

EXPERIENCE

Data Scientist | Atriano | Remote, CA   September 2024 – Present
- Building a generative storytelling web app for an **EdTech** client using the *LLaMA 3.2* (*LLM*), fine-tuning with *LoRA* (*Python, Hugging Face & AWS*) by utilizing a custom dataset, and integrating evaluation for toxicity filtering & age-appropriate vocabulary.
- Engineering a collaborative filtering *recommendation system* to suggest personalized story topics based on user demographics.

Data Engineer, Imaging | ImagoAI | Milpitas, CA   October 2024 – January 2025
- Deployed Galaxy (*Food Quality Testing*) for *10 clients* by fine-tuning ML models (*Python*) on client samples hyperspectral images.
- Automated (*Python* & *Shell*) the process of Hyperspectral Image camera calibration *increasing the process efficiency by 80%*, *reducing the manual efforts by 95%*, and *saving $30000*.

Data Scientist | Quinbay | Bengaluru, India   May 2022 – October 2022
- Spearheaded deployments (*E-commerce*) using *Python*, *PyTorch*, *Node*, *React*, & *SQL* and analytics with *GCP* & *Tableau*.
- Developed and optimized quantized object detection model to automate order return approvals for electronic products *reducing processing time from 6 hours to 15 seconds*.
- Developed a pipeline to flag counterfeit products (*key brands like Apple & Samsung*) leveraging *Computer Vision* & *Natural Language Processing* for product screening. This feature *reduced manual efforts by 30%* and *accelerated the process by 70%*.

Machine Learning Engineer | Hono | Gurugram, India   July 2021 – April 2022
- Deployed Predictive Analytics pipelines (*Django* & *JavaScript*) to automate production of client specific services (*HR Tech*) using ETL/ELT pipelines (*Python*, *AWS*, *SQL*), Machine Learning (*Python*), & A/B Testing (*VWO*) *generating $120000+ revenue in 2022*.
- Developed *Employee Attrition* (*estimate resignation chances*), *Promotion Suggestion* (*recommend employees for promotions*), *Hike Estimation* (*suggest salary hikes per employee*), & *Leave Prediction* (*predict employees leaves for 30 days*) features for *12 clients*.
- Optimized data infrastructure to *reduce model update latency from 45 to 10 minutes* by leading ETL *redesign* and *cross-functional collaboration* with backend and data warehousing teams.

Data Scientist- Intern | ImagoAI | Gurugam, India   April 2021 – May 2021
- Designed Galaxy tests (*Food Quality Testing*) for *3 products* (Corn, DDGS & Wheat) using ML Algorithms (*Python* & *AWS*).

Research Intern | Department of Physics, IIT Roorkee | Roorkee, India   September 2020 – June 2021
- Experimented on Deep Learning models (*TensorFlow*) for semantic segmentation to estimate ocular tumor growth and trends.
- Developed an image annotation tool (*MATLAB*) *reducing annotation time by 85–90%* (from 30 to 3 days), eliminating dependence on external services, and enabling faster & more accurate model training across multiple ongoing projects of the lab.

SKILLS

Programming Languages: Python, C, C++, SQL, JavaScript, Java, C#, HTML, CSS, R, MATLAB, Perl
Data Science Skills: PyTorch, TensorFlow, Keras, Tableau, PowerBI, PySpark, Databricks, Qlik, Snowflake, Hadoop
Software Skills: Node, React, MongoDB, Django, Kafka, Jenkins, Selenium, Git, Kubernetes, Docker, AWS, GCP, Jira, Confluence

PROJECTS

Metaheuristic Optimization vs Backpropagation | Computer Vision, University at Buffalo   January 2024 – May 2024
- Developed (*PyTorch*) Metaheuristic optimizers using Particle Swarm Optimizer and Grey Wolf Optimizer for Convolutional Neural Networks training and contrasted their performance with backpropagation (Stochastic Gradient Descent and Adam Optimizer).

Top Spotify Tracks Database | Data Models Query Language, University at Buffalo   August 2023 – December 2023
- Built a web application (*Flask app*) to showcase an intuitive dashboard for "Top 50" Spotify playlists across 67 countries.
- Engineered a pipeline (*Python*) to aggregate real-time data from Spotify Web API in a normalized relational database (*SQL*).