Abhishek Kumar
Senior Data Scientist

+91-9962645223
Gurgaon, Haryana
reachme.abhishek.kr@gmail.com

Portfolio: abhinine4.github.io
github.com/abhinine4
linkedin.com/in/akumar58

Senior Data Scientist with 5+ years of experience in machine learning, computer vision, and NLP, with a Master's in Computer Science & Engineering (University at Buffalo). Proficient in Python, C++, and PyTorch. Relevant coursework: Machine Learning, Deep Learning, Computer Vision, Information retrieval and Distributed Systems. I'm currently working on a generative AI based conversational chatbot for personalized financial insights. I'm Seeking an Applied Scientist role to solve real-world problems.

SKILLS

Language and Database	Python, C++, SQL, Java (familiar), MongoDB, Weaviate, Milvus, MySql, Redis, ElasticSearch
Tools and Monitoring	Git, Docker, Azure, AWS, Jenkins, New Relic, Grafana, Prometheus, Swagger
Framework and Libraries	PyTorch, TensorFlow, Langchain, Django, FastAPI, Transformers, OpenCV, Sklearn, NumPy, Kafka

WORK EXPERIENCE

Senior Data Scientist (Lens.Ai, Mbk-Parser, Help-Gen-Bot)　　　　　　　　　FEB 2024 — Present
Mobikwik　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　*Gurgaon, Haryana*

- Developed **Lens.Ai**, a generative AI-powered financial advisory chatbot using **NLP, machine learning** and **Django**, delivering personalized tax and financial advice based on user earning and spending patterns.
- Developed **text-to-NoSQL** query generation module using LLMs, processed 200+ unique categories/subcategories, payment methods, merchants and associated banks.
- Designed a vector database (Weaviate) **caching system** with hybrid matching, cutting system costs by 80% for frequent queries.
- Engineered multi-turn conversation flows with context management using **Langchain**.
- Created a natural language generation module integrating user data and conversation history for human-like responses.
- Launched **lens-test-suite**, an automated validation tool to ensure consistency in chatbot performance across model updates.
- Integrated **Prometheus** and **Grafana** for real-time monitoring of ML APIs and system health.
- Built **MbK-Parser**, a scalable credit card statement parser using Microsoft **Azure** OCR and **OpenAI** GPT APIs, processing high-throughput data with **Kafka** and **FastAPI**.
- Fine-tuned open-source **LLMs** (Llama, Qwen) with **PEFT** on mobikwik SOPs and added **RAG** pipelines for **Help-Gen-Bot**, resolving customer queries with 85%+ accuracy.

Research Engineer (SemaFor)　　　　　　　　　　　　　　　　　　　　SEP 2022 — FEB 2024
Artificial Intelligence Innovation Lab (A2IL)　　　　　　　　　　　　　　*Buffalo, New York*

- Developed a **text+pose**-guided image morphing **computer vision** system using latent diffusion models (ControlNet, Stable Diffusion), achieving high-resolution outputs.
- Implemented latent interpolation and **CLIP**-based scoring to optimize image generation quality.
- Developed image manipulation detection system to **detect, localize, and label** tampered parts in news articles. Masters Thesis
- Built image manipulation detection system with a **ResNet-50**-based CNN, to detect JPEG compression errors with 93% accuracy.
- Utilized **NEDB-Net** to extract noise and edge-based features to localize manipulations for the tampered regions.
- Fine-tuned a custom **Yolo-v8** model to detect objects in the localized regions and classify them into 18 categories.
- Created a **NLP** transformer tool with **SpaCy** for controlled named entity and parts of speech replacements in news articles.

Systems Engineer (Cummins)　　　　　　　　　　　　　　　　　　　　FEB 2018 — AUG 2020
Infosys　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　*Bhubaneswar, Odisha*

- Trained a **machine learning** model (Random Forest) to classify refrigeration system components, achieving 97% accuracy.
- Built an **XGBoost**-based price prediction module, integrated into Cummins' pricing system.
- Automated batch job monitoring with **RPA**, reducing data transfer failures by 20%.
- Managed CI/CD pipelines with **Jenkins** and **Git** for system reliability.

EDUCATION

Master of Science in Computer Science & Engineering, *University at Buffalo, New York, GPA - 3.72/4*　　AUG 2023
Bachelor of Technology in Mechanical Engineering, *SRM University, Chennai, CGPA - 8.8/10*　　　　　MAY 2017

PROJECTS

Soccer Player Re-Identification
- Developed a deep learning system to re-identify soccer players in broadcast videos, achieving 63% mAP.
- Trained dual-branch network with ResNet and OpenPose backbones, using PyTorch, OpenCV, and Triplet Loss.
- Applied bilinear pooling to fuse appearance and pose features, enhancing robustness in dynamic scenes. Technical Report

Ear Hair-Cell Detection and Counting
- Built a computer vision pipeline to detect and quantify damaged ear hair cells for deafness studies in animals.
- Utilized Python, OpenCV, and Hough Transform for feature extraction, with DBSCAN clustering for cell segmentation.
- Implemented template matching and non-max suppression to improve detection precision, supporting biomedical research.