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

Abhishek Kumar

Senior Data Scientist

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
Framework and Libraries	PyTorch, TensorFlow, Langchain, Django, FastAPI, Transformers, OpenCV, Sklearn, NumPy, Kafka, vLLM
Tools and Monitoring	Git, Docker, Azure, AWS, Jenkins, New Relic, Grafana, Prometheus

Work Experience

Senior Data Scientist **FEB 2024 — Present**
Mobikiwk *Gurgaon, Haryana*

- **Lens.AI**, Developed 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, which processes 200+ unique categories/subcategories, payment methods, merchants and associated banks to create accurate NoSQL queries.
- Designed a vector-based **caching system** (Weaviate) with **RAG** and hybrid search, reducing latency by 70-80% and costs by 90%.
- 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.
- **Bank-Statement-Parser**, Developed a scalable credit card statement parsing service using OCR and large language models, deployed on **AWS, Kafka, FastAPI, mongoDB**, processing high-throughput data for millions of customers.
- The parsing service uses Azure OCR to extract text from incoming statements and Open AI's Gpt-4o-mini for response formatting.
- **Help-Gen-Bot**, Fine-tuned open-source LLMs (Llama, Qwen) using **PEFT(QLoRA)** on custom SOPs, integrating RAG pipelines to resolve customer queries, achieving a 20% faster resolution rate.

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

- **Face-Morphing**, 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.
- **Manipulation-Detection**, Developed manipulation detection system to **detect, localize, and label** tampered news articles.
- Trained a CNN model using differential images with **ResNet-50** as backbone, to detect compression artifacts 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 label them into 18 categories.
- Engineered text-transformer tool for controlled entity and parts of speech replacements using **SpaCy** and **transformer (BERT)** models, enhancing data preprocessing for downstream AI tasks. Used **co-reference** resolution to maintain context in long texts.

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

- Trained a **Random Forest** model to classify refrigeration system components for CUMMINS, 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*, Masters Thesis — 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

CNN-VAE Face Image Generation

- Developed a CNN-VAE model in PyTorch to generate realistic face images from a custom face dataset.
- Designed a 4-layer encoder-decoder with a 256D latent space, achieving a validation loss of 160.03 using MSE and KL divergence.