+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, Celery, Kafka, vLLM
Agents and Tools	Crew.ai, Langraph, Google-ADK, Git, Docker, Azure, AWS, Jenkins, New Relic, Grafana, Prometheus

Work Experience

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

- **Vaarta-Vedh**, 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 AWS textract to extract text from pdf files and LLMs for parsing, validation and response formatting.
- **Help-Gen-Bot**, Fine-tuned Llama3.2 3b model using **PEFT(QLoRA)** on custom SOPs and integrated RAG pipelines to resolve merchant transaction queries, achieving 20% faster resolution rate. Served fine-tuned models using **vLLM** on AWS EC2 instances.

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 independent and dependent system components with 97% accuracy.
- Built a part price prediction module using **linear regression** for refrigeration units, supporting faster data-driven quotations.
- Managed data workflows and CI/CD pipelines using **Jenkins, Git, Airflow, and PySpark**, ensuring reliable and automated deployment of pricing models and data pipelines.

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 model using Pytorch to re-identify soccer players in broadcast videos, achieving 63% mAP.
- Trained dual-branch network with ResNet and OpenPose subnetwork as backbones to learn appearance and body part features.
- Applied bilinear pooling to fuse appearance and body part features, enhancing robustness in dynamic scenes. Technical Report

CNN-VAE Face Image Generation

- Built a CNN based Variational Autoencoder (CNN-VAE) in PyTorch to generate realistic face images, leveraging a custom dataset and latent space representations to learn meaningful facial features.