+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

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*

- **LENS.AI :** Developed a **conversational financial advisory chatbot** using generative AI that provides personalized financial and tax advisory based on users spending, earning and investment patterns.
- Developed **text-to-mongo (NoSQL)** query generation module using LLMs to fetch relevant user data. The module processed 200+ unique categories/subcategories, payment methods, merchants and associated banks to create custom NoSQL queries.
- Created a **query caching** module to cache templated queries and FAQs. Used vector databases (Weaviate db) and **hybrid matching with re-ranking** strategies to fetch common user queries. Reduced system latency and cost by 70-80%.
- Developed **custom routing methods** and **conversation context management** to manage multi-topic and multi-turn chat flows.
- Built a **natural language generation module** that uses conversation history, user query and custom user data to create human-readable responses.
- Created lens-test-suit, a fully automated testing and validation service that tracks model and prompt changes to maintain consistency in chatbot responses.
- Integrated **Grafana, New Relic, and Prometheus** in ML services for logging and monitoring API requests and responses.
- **MBK-PARSER :** Built a scalable **credit card statement parsing service**, that extracts texts from credit card statements and parses them to notify Mobikwik users for bill payments.
- Utilized Azure OCR APIs and large language models to parse extracted key-values pairs and transactional data.
- Scaled parsing service using **Apache Kafka, AsyncIO and Multithreading with thread pool executor** to handle high throughputs.
- Used Python, AWS S3, MongoDB, Azure AI services, AsyncIO, Fast API, Kafka, OpenAI to create the parser service.
- **HELP-GEN-BOT : Fine tuned open source LLMs** (Llama and Qwen) on custom SOPs to handle customer queries.
- Utilized **PEFT with QLoRA**(4bit) from bitsandbytes and Unsloth library to efficiently fine-tune LLM models on custom data.
- Used **RAG** pipeline to fetch relevant metadata and FAQs to aid llm prompts in issue resolution.

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

- Created a **text+pose guided image morphing** system using latent diffusion models (ControlNet + Stable diffusion weights).
- Added noise and performed latent interpolation on best pair human face images to generate high-resolution morphs.
- Generated images at different noise levels and selected the best candidate by computing CLIP similarity scores for fixed prompts.
- Developed an **image manipulation detection system to detect, localize, and label tampered parts in news articles**. Thesis
- **Trained a CNN model with a Resnet-50 backbone to detect jpeg compression errors** in tampered images 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.
- Developed a text-transformer tool to perform controlled named entity and parts of speech replacements in texts using SpaCy.

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

- Trained **random forest model to classify dependent and independent parts** for refrigeration systems with 97% accuracy.
- Developed a module to **predict part prices using XGBoost regressor** and integrated with CUMMINS part pricing system.
- Developed employee task delegation page to manage production workflows and approvals using python.
- Reduced batch data transfer failure rate by 20% through **automation and monitoring of batch jobs** using RPA and Appworx.
- Managed deployment and maintenance pipelines for system availability and reliability using Jenkins and Git/Github.

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

Player Re-Identification: Trained a dual branch network with appearance and body part features extracted from ResNet and OpenPose subnetwork as backbones, to re-identify soccer players in broadcast videos with 63% mAP. Used Python, PyTorch, OpenCV, Bilinear Pooling, Triplet Loss. Paper

Ear Hair-Cell Detection and Counting: Detected and extracted damaged inner and outer ear hair cells to measure deafness in animals. Used Python, Template Matching, Non-max Suppression, DBSCAN Clustering, Houghs Transform.