

Indrajeet Kumar

Profile snapshot

Senior Data Scientist with 6+ years of experience in building end-to-end ML systems, specialized in GenAI, computer vision, deep learning, and statistical modeling. Proven expertise in scalable ML pipelines, microservices-based architectures, and production deployment using FastAPI, Docker, and cloud-native tools. Experienced in agentic systems, model serving, and real-time analytics.

Education

BTech & MTech (IDD)	IIT BHU Varanasi	2014-19	7.05
---------------------	------------------	---------	------

Work Experience

Bosch (Senior Data Scientist) @ Bangalore	2023 - current
--	-----------------------

Worked for developing high-throughput scalable data curation services in the data delivery domain to identify and extract diverse and long-tail edge cases data for improving robustness of automatic driving assistant system(ADAS) perception models for dynamic objects(vehicle, pedestrian, car, large vehicle) use cases.

Designed and developed an agentic service for automated video label quality control, processing millions of records at scale: [Autolabel Quality Check].

- Designed and implemented a multithreaded, queue-driven pipeline for efficient label ingestion, processing, and large-scale error analysis.
- Built an agentic validation framework using LangChain to detect false negatives (FN) and false positives (FP) via statistical and deterministic rule-based checks.
- Integrated Gemini-2.0-flash-light as the foundation model for frame-level analysis and validation workflows.
- Containerized and deployed services using Docker with scheduled execution via cron-based job orchestration.

Tools & Technologies: Databricks, multithreading, langchain, LLM, cron-services, docker, queue.

Designed and implemented a pipeline for detecting and extracting salient temporal segments from long-form video data: [video snippet selection].

- Designed and implemented a scalable ML pipeline for extracting salient timestamps for dynamic objects(car, vehicle, large vehicle) in long-duration videos.
- Performed time-series analysis on video frame embeddings generated by DINOv2 to understand temporal dynamics and scene transitions.
- Conducted statistical analysis on time-series data to detect temporal discontinuities, irregular intervals, and structural breakpoints.
- Applied local variance and signal-based methods for robust detection of gaps, anomalies, and regime transitions.
- Built and deployed scalable RESTful APIs for data retrieval and model inference within a microservices architecture.
- Optimized video data curation pipelines to extract diverse and high-quality snippets, enhancing training efficiency for transformer-based perception and planning systems.

Tools & Technologies: FastAPI, REST APIs, Dinov2, video processing, Kalman filter, Docker.

Design and Build Application for Image Retrieval using Scene Graph Generation: [SIGNIFI]

- Designed and built an end-to-end ML microservices platform for image retrieval using scene graph generation.

- Developed multiple FastAPI-based services for model inference, embedding generation, and metadata storage.
- Designed ontology and schema for scene graph storage in MongoDB.
- Containerized services using Docker and deployed as independent microservices components.
- This helped in retrieving diverse data selection and specific scenarios based on edge cases for robust model training and improvement.

Tools & Technologies: Docker, FastAPI, MongoDB, image embedding, PyTorch.

Innominds (Ace Vision Group) (Data Scientist) @ Bangalore	2021 - 2023
--	--------------------

- Worked on image processing for OCT images of the eye. Finding eye pore diameter, distance, volume, radius, depth using detection and segmentation technique.
- Created hybrid network using U-Net and Mobile Net arch to get real time inference.
- Fine-tune YOLO-v5 model for eye pore detection.
- Implemented custom weighted BCE + Dice loss for segmentation.
- Implemented specific automatic annotation technique using image processing.
- Used multiprocessing and multithreading to run applications faster.

Tools & Technologies: Python, Pytorch, Computer vision, Azure, cv2.

Boltzmann Labs (AI Developer) @ Bangalore	2019 – 2020
--	--------------------

Worked on building AI Libraries: Full-Stack Drug Discovery pipeline using AI.

- Developed the model for Drug-Target binding affinity prediction using CNN.
- Developed GCN models for drug properties prediction.
- Developed GCN based Multi label classification model for metabolic pathway prediction.
- Implemented interpretabilities for all the models using ex-Integrated Grad, LRP.

Tools & Technologies: Python, Rdkit, Graph Convolution Network, Graph attention network, Convolution Neural Network.

Technical Exposure:

Language: Python, C++ | **Database:** Databricks, Mysql, MongoDB | **ML Tools:** Sklearn, Scipy, Numpy, statsmodels | **DL Framework:** PyTorch | **ML Algorithms:** KNN, Logistic reg, SVM, Random Forest | **DL Algorithms:** MLP, CNN, RNN, GCN, Transformer | **Application development:** FastApi, Docker, Kubernetes | **GEN AI:** langchain, langgraph, MCP, Huggingface, RAG.

Online Courses:

Applied Machine Learning from Applied AI.

Position Of Responsibility

Training and Placement Head coordinator 2018-19.

Analytical event head Spirit-17.

Event coordinator of dakshana at IIT BHU.

Links

 <https://github.com/abchotujnn1/>
 <https://www.linkedin.com/in/indrajeet-kumar-654aa1148/>
 <https://medium.com/@indrajeetkumar.iitbhu14>
<https://abchotujnn1.github.io/indrajeet.github.io/>