Ankit Pal

+1 (669) 240 0188 | ankitpal3398@gmail.com | github.com/apal9569 | linkedin.com/in/ankit9569/

EXPERIENCE

Research Assistant, *San Jose State University (Autonomous Systems)* | (California, USA) Oct 2023 - Present

- Implemented **object detection** and **segmentation models** using **YOLOv11** and **Faster R-CNN** to enhance real-time robot navigation capabilities.
- Integrated **Kalman Filters** and **Extended Kalman Filters** for sensor fusion, significantly improving **object tracking** and **localization** in dynamic environments.
- Built scalable image classification pipelines using **Vision Transformers** and **Swin Transformers**, significantly enhancing feature extraction and model performance for high-resolution imagery.
- Utilized **PyTorch, TensorFlow, and MLFlow** to develop scalable deep learning pipelines, improving deployment efficiency by 40%.

Data Scientist, *Enliant Inc* | Remote (Texas, USA) May 2023 - Aug 2023

- Developed a computer vision-based quality assurance system for **e-commerce product image validation** using **EfficientDet** and **DETR** models, achieving 97% accuracy in identifying defective or non-compliant product images.
- Designed and deployed end-to-end machine learning pipelines on **Google Cloud Vertex AI** integrated with **CUDA-optimized TensorRT**, ensuring low-latency inference for real-time production systems.
- Designed **VAEs** and **implemented GAN-based models** for synthetic data generation, reducing dependency on real-world datasets.
- Utilized **OpenCV** and **Albumentations** for advanced preprocessing and data augmentation, improving model robustness against noisy inputs.

Data Analyst, *Optum - UnitedHealth Group* | India Aug 2020 - July 2022

- Developed and improved **SAS**, **Java**, and **Scala** analytics for insurance claims processing, collaborating with business and operations teams to define **key performance indicators (KPIs)**.
- Established data governance across **SQL and cloud source (Azure)**, reducing data failures by 30% and report time by 20% through streamlined ETL workflows in Python and Scala.
- Transformed the data as per business rules. It brought unification and normalization in data using optimized **SQL queries** and **database indexing** and provided Key KPIs to make crucial business decisions.
- Performed CI/CD using **Azure DevOps** for build and release operations, utilizing **Azure Repos** as the version control system.
- Employed **Change Data Capture**, **Slowly Changing Dimensions** for auditing full/partial loads, and recovery from erroneous data and implemented **MD5 hashing** for record deduplication and data profiling using SQL Stored procedures to establish a 'chain of trust.'
- Engineered 200 batch and stream **ETL pipeline** workflows from diverse value chains and data formats into **ADLS**, **Implemented Airflow** for dataset onboarding, and utilized Spark jobs, resulting in a 55% reduction in ingestion time.
- Utilized **Kafka** pub-sub model for tracking real-time events in the data records to trigger processes for data orchestration.
- Orchestrated **Terraform (Infrastructure as Code)** scripts resulting in a 40% reduction in provisioning time.

SKILLS

Languages	Python, C/C++, Java, Scala, Javascript, SQL, PySpark
Databases	Relational (MySQL, SQL Server, PostgreSQL), NoSQL (MongoDB, Cassandra), Snowflake
Cloud	AWS, Microsoft Azure, Azure Data Factory, Microsoft SQL Server, AWS S3, Kubernetes
DL Framework	PyTorch, TensorFlow, Keras, Scikit-learn
Big Data Tools	Apache Kafka, Spark, GCP (BigQuery), Airflow, Hadoop, Hive
ML Tools	MLFlow, Kubeflow, Google Cloud Vertex AI, TensorRT
Computer Vision	OpenCV, EfficientDet, DETR, Transformers, DeepLabV3+, DensePose, ResNet, MobileNet

EDUCATION

San Jose State University, *MS in Artificial Intelligence* | California, USA GPA: 3.63 / 4.0 Dec 2024

National Institute of Technology, *B.Tech in Electrics and Communication* | Delhi, India GPA: 7.49 / 10 Aug 2020

PUBLICATION

Algorithm for Distracted Driver Detection and Alert Using Deep Learning | Springer July 2021

- Model divides driving behavior into ten categories, nine dealing with drivers distracted by other activities and one with "safe driving". Our system will alert the driver if he is determined to be in any of the nine distracted classes, reducing the likelihood of an accident.