RISHI RAJ SINGH

coderishi9@gmail.com | +91 8369479713 | www.linkedin.com/in/coderishi9 | https://github.com/coderishi9

PROFESSIONAL EXPERIENCE

KlearStack | AI driven OCR and Intelligent Data Extraction Software — Pune, India
Backend Machine Learning Engineer — Oct 2021 - Present

- Leveraged open source LLM models to improve the prediction pipeline for different document types. Improving the overall prediction from existing 82% to 94%. Used Quantization techniques and fine-tuned the large models to run on affordable compute resources reducing expenses 4 times.
- Researched and deployed end to end LLM inference pipeline using vLLM, tensorRT-LLM, fastAPI and Redis.
- Developed Auto document type classification pipeline for the product leveraging state of the art transformers models and set up scalable inference engine as a consumable REST API. Accuracy improved from 64% to 97% as compared to the existing approach.
- Implemented Translation pipelines to handle multilingual documents helping the product to enter into markets globally without language being a barrier.
- Developing and Maintaining Core Product for Line Item Extraction from Documents using Computer Vision, Pattern Recognition and NLP

Innefu Labs | Image and Video Forensic Software — Delhi, India
Data Scientist — May 2021 – Sep 2021

- Researched on various scarcely pursued Image processing techniques such as Non-Linear debarring, 3D measurement in a 2D image, Perspective Stabilization etc.
- Developed Computer Vision Pipeline for Harmful items detection like Guns, knives, Screwdriver, etc. in Airport X-Ray dataset using MMDetection toolbox leveraging Computer Vision models like Yolo, RCNN, etc. and transfer learning achieving an impressive mAP@[0.5:0.95] of 0.63.

Buddihealth | AI Based Software to Convert Unstructured EMR into Structured Data — Chennai, India
Machine Learning Intern — Jul 2020 – Apr 2021

- Applying various Machine Learning techniques on Medical Data and Electronic medical records(EMR) to develop various API and web-apps.
- Research with various state-of-the art algorithms like Bi-LSTM-crf, CNN-LSTM etc. for named-entity recognition, relationship extraction etc.

EDUCATION

Lovely Professional University — Jalandhar, India
Bachelor of Technology, Computer Science & Engineering — May 2021
Cumulative GPA: 8.34 /10.0

PROJECTS

Named-Entity Recognition Using Bi-LSTM-CRF Model — 2021

- Bi-LSTM-CRF is a state of the art approach for named-entity recognition.
- Model consisted of 2 LSTM layers one from front and one from Back, an embedding layer and a CRF layer on top for pattern recognition.

DeepFake detection Using CNN-LSTM Model — 2022

- CNN-LSTM is a dual architecture model with LSTM connected to a CNN so that the model can remember information about previous frames of video.
- The model was able to recognise the fake parts with 98.67% accuracy.

SKILLS

- Technical Skills: Python, Django, Flask, FastAPI, SQL, NoSQL, MongoDB, Redis, Docker, Git, Github
- AI/ML frameworks: Scikit-Learn, Keras, Tensorflow,, Pytorch, Spacy, NLTK, Hugging Face, OpenAI APIs
- Machine Learning (ML), Deep Learning(DL), Artificial Intelligence (AI), Natural Language Processing
- Environments and Frameworks: Azure ML, AWS Sagemake