Divya Kuwarbi

divyakuwarbi@gmail.com | 8920807811 | Delhi NCR | Linkedin.com/in/divya-kuwarbi-a7608019b | Github-github.com/divya-kuwarbi17

SUMMARY

Data scientist with nearly 3 years of experience in building machine learning and deep learning models in the realms of Computer Vision and NLP, capable of building, deploying and testing models as Flask API.

EXPERIENCE

AI Research Engineer, **Acxiom technologies**, Andheri east, Mumbai *Dec 2023 - Present*

Sales ChatBot/FAQs ChatBot
- Integrated **GPT-3.5-turbo-16k** for resolving user queries with high precision and conversational fluency.
- Enhanced lead generation by crafting engaging and informative responses tailored to user needs.
- Implemented a vector database(**Milvus DB**) to store and retrieve embeddings for improved context retention and query matching.

Indian KYC Document Data Extraction System. [Extracts Data from kyc's (Aadhar, Pan, voter, passport etc)]
- Fully offline, on-premises KYC Document Data Extraction System deployed on Indias largest bank server (AWS), ensuring complete data security with no external dependencies.
- Developed an KYC Detection using YOLOv5 and custom CNN trained on 18K+ images at 1280 resolution, achieving 98% accuracy on production data
- Developed a robust classification system for Driving Licenses, addressing the countless formats across Indian states by using K-Means clustering for grouping similar formats and YOLO for precise field detection. with 95% accuracy for classification

Car Inspection System
- Designed and utilized synthetic datasets to simulate real-world scenarios, ensuring robust model performance across diverse conditions
- Accurately identified vehicle attributes such as make, model, color through trained classifiers and image analysis.
- Trained models to recognize car views (e.g., front-left, back-right) with precision.
- Successfully implemented algorithms to identify and classify car parts (e.g., hood, bumper, doors, fenders) using segmentation techniques.

Data Scientist, **Althea.ai** , Noida *Jan 2022 - Dec 2023*

Document Digitization:
- Developed and deployed Image Classification using transfer learning, multimodal gated fusion network
- OCR using Textract, **pytesseract**, KYC Entity Extraction using **LayoutLM**, **spacy,** sci-spacy
- Preprocessed Images using Opencv, numpy, Object Detection using YOLOv5, YOLOv8

Fraud Detection Insurance Claim:
- Applied Error Level Analysis (ELA) to identify image manipulations, such as alterations or forgeries, with high precision. Utilized cosine similarity to detect duplicate submissions and flag repeated images across claims efficiently.
- Automated the detection and verification of document stamps and signatures to uncover inconsistencies and potential forgeries.
- Designed a robust framework of fraud triggers based on extracted entities and their relationships to detect anomalies.

Video KYC:
- Developed a next-generation Video KYC system to streamline identity verification by combining object detection with YOLOv5 and automated entity extraction using regex.
- Designed to detect user presence and extract critical details like ID numbers and names directly from documents

Hospitalization Risk
- EDA, Feature Selection and model selection
- Working on FHIR data for predicting hospitalization risk for heart patients.

Deployment:
- Got a touch of **RabbitMQ**, **Consumer building**, **git commands**.

EDUCATION

M.Sc. Computer Science, Big Data Analytics *Nov 2020 - Jun 2022*
Central University of Rajasthan GPA: **9.36**
Bachelors of Science, Computer Science *Jul 2017 - Oct 2020*
University of Delhi GPA: **7.5**

Intermediate, Science(Non Medical) *Apr 2016 - May 2017*
GM Sr. Sec. School Percentage: 86.6

SKILLS

Python3: pandas, numpy, scipy, sklearn, matplotlib, scikit learn, NLTK
Machine Learning: Linear Regression, Logistic Regression, SVM, KNN, Naive Bayes, Random Forest,XGBoost
Deep Learning: CNN, RNN, LSTM, Transfer Learning Models such as VGG16, Rsenet50 etc., YOLOv5,SAM for segmentation and object detection, OCR-easyocr, pytesseract, paddleocr, Tensorflow, Pytorch
Data Mining: Data Preprocessing, Data Cleaning
NLP: Spacy, SciSpacy, BERT, Attention, TFIDF, LLM (llama2 7b chat, Mistral-7B-Instruct-v0.1), GPT3.5t16k Sentence Transformer models, Donut, Hugging face Transformers

Certification

Deep Learning Certification, , **NPTEL-Swayam**

PUBLICATION

Enhancing Generalizability in Biomedical Entity Recognition: Self-Attention PCA-CLS Model (IEEE)

One of the primary tasks in the early stages of data mining involves the identification of entities from biomedical corpora. To address out-of-vocabulary challenges in biomedical text, the PCA-CLS (Position and Contextual Attention with CNN-LSTM-Softmax) model combines global self-attention and character-level convolutional neural network techniques.