

Omjee Yadav

📍 Greater Noida ✉ yadavom2000@gmail.com ☎ 7007027984 🔗 LinkedIn

TECHNICAL SKILLS

Programming Languages: Java, Javascript, Python, SQL

Concepts: Full Stack Web Dev, Generative AI, DBMS, Cloud, Data Structures and Algorithms, OOPS/ System Design

Web Frameworks: React, NodeJs, Spring/Springboot, FastAPI.

Developer Tools: Cloud, Figma (UI/UX), Docker

PROFESSIONAL EXPERIENCE

COFORGE, Software Engineer

06/2023 – present | Greater Noida, India

LLM based Data Extraction from Automotive Repairs Orders

- Developed a novel algorithm for **Restructuring OCR** data into actual document layout, which provides **positional context** to **LLM** for inference leading to a **30% increase** in accuracy compared to raw OCR data.
- Employed **NLP**-based techniques and bounding boxes for **classifying** and removing irrelevant sections from documents **saving \$100k+** in project operating costs by reducing the input tokens.
- Achieved a **50% reduction** in models output response time by **parallelization** of LLM pipeline, decreasing output tokens using **Prompt Engineering** techniques and reformatting them in post-processing

Automobile Parts Logistics Tracking System:

- Developed **Spring Boot** backend with **MVC** architecture for a logistics application managing **packing, vanning, and invoicing**
- Implemented **APIs** and reports using **JRXMLs** to facilitate data analysis and streamline production/export workflows.
- Utilized **JUnit** for robust backend testing and ensured data integrity through **MySQL** database management.

WebApp for Document Processing

- Created Frontend using **ReactJS** for batch upload of documents for data extraction, annotation and editing
- Developed **Backend** using **FastAPI** to extract fields from uploaded documents using Layout based models like Azure **Form Recognizer** and **finetuned** NLP models (spacy) and extracting data using **bounding box** algorithms.
- Displayed extracted fields and in tabular format and provided option for manual annotation and validation of extracted data.

Advanced Document Querying using RAG:

- Created Backend using **FastAPI** for **document classification** and then **extracted** the fields based on the type of documents.
- Converted PDFs into text data using custom **OCR** pipeline which included **handwritten** text detection and **layout** detection
- Generated **embeddings** from text and stored them in a **vector database** (FAISS,pinecone) for fast querying
- Used Llama 2 LLM model with **RAG (retrieval augement generation)** for document querying, made pipeline using **LangChain**

Custom OCR Model Development:

- Developed a **CNN** based model to detect **rotation** and **skew** in images. trained on custom dataset and did **augmentation**
- Used **docTr** library for bounding box creation ,which uses **VGG16** and **resnet50** model
- Created **custom algorithm** for transforming raw OCR data into structured text markup and layouts, enhancing readability
- Implemeted **table layout detection** and converted the data into an editable csv file to be used in **excel**.

PROJECTS

Disease Spread Prediction using Sentiment Analysis of Tweets

02/2023

- Performed **Sentiment Analysis** and Multi Class Disease Classification on PHM2017 benchmark dataset using **NLPK**
- Fetched tweet text from tweet IDs using **SNScrape**, removed links, hashtags and **stopwords**, performed **lemmatization** and **vectorization** for preprocessing
- Did comparative analysis of various algorthims on this using random forest,decision trees,multinomial Naive Bayes etc.

Blogging Website using Spring Boot

- Developed a robust blogging platform using **Spring Boot** as backend. Implemented user roles and permissions with **Spring Security** and **JWT**, allowing admins to manage categories and posts while ensuring secure user **authentication**

EDUCATION

Galgotias College Of Engineering And Technology,
B.Tech(Computer Science Engg.)

2019 – 2023 | Greater Noida