

MD SUHAIL AHMED

<https://www.linkedin.com/in/md-suhail-ahmed-4578a125b/>

[✉ mdsohailahmed1843@gmail.com](mailto:mdsohailahmed1843@gmail.com)



<https://github.com/Sohail1173>



6362048485

Profile: I Passionate about Data Science & AI, having good object oriented programming skills, well versed knowledge of making full end-to-end Data Science pipeline from Data Gathering to Scalable deploying applications in various cloud platforms, having good analytical and problem solving skills.

Work Experience

1. AI Engineer Intern

- **Algorithmia digitech pvt ltd: Apr 2023-July-2023**
- **FinGPT: A Large Language Model.**
- Collaborated with a team to develop the **SmartBasket.ai** mobile application.
- Contributed to the design and implementation of key features.
- Preprocessed and cleaned the collected data to ensure high-quality inputs for the model.
- Trained the LLM on a diverse dataset of company information, allowing it to generate relevant and contextually accurate text.
- Fine-tuned the model using **custom langchain agents**, optimizing its performance for specific business applications and use cases.
- Participated in regular team meetings and provided updates on progress.
- **Technology Stack: Langchain, Elasticsearch, BeautifulSoup, Huggingface, Transformers.**

Personal Project

1. Chat With Multiple PDF'S:

- This project utilizes the **large language models**.
- This web app allows you to upload multiple PDF files and QnA from within them using **OpenAI API Key**.
- The first step of this project involves extracting all the pages from the uploaded PDF files and storing them in a db. To ensure that the OpenAI limits are not exceeded, a **Recursive Text Splitter** is employed to divide these pages into smaller manageable chunks. This division enables efficient processing of the content while staying within the constraints of the **OpenAI API**.
- After the pages are extracted and divided into smaller chunks using the Recursive Text Splitter, the next step in the project is to convert these chunks into **OpenAI embeddings**. **OpenAI embeddings are numerical representations of text data** that capture its semantic meaning. By converting the chunks into embeddings, the web app can effectively analyze and process the content in a way that can be understood and utilized by the large language models. These embeddings are crucial for various natural language processing tasks, including **question-answering and semantic search**, which are the core functionalities of the application. Once the conversion is completed, the embeddings are then stored in the **'chromadb' vector store**, making them easily accessible for subsequent processing and querying
- **Tech Used: Langchain, OpenAI Embeddings, Chromadb, Tiktoken, Streamlit**
- https://github.com/Sohail1173/Chat_with_files

2. Image Recommendation

AIM: i build a web app in which user can search the desired item of image by putting their image. Then my CNN & ML models will find out the 5 similar images from my database. which i prepared for this project. in which i collected 8k images of different class of Animals items.

Tech Used: CNN, ResNet50, Nearest Neighbour, Tensorflow, openCV, Image processing.

- <https://github.com/Sohail1173/Reverse-Image-Search>

Core Skills

Languages

Python, Elasticsearch, Object Oriented Programming.

Deployment

Docker

Google Container Registry
Amazon Container Registry
CloudRun ,

Basic Data Structure

Linear Search, Binary Search,
Bubble Sort, Quick Sort,
Insertion Sort, Selection Sort,
DataFrames, Series, Hashing

Mathematics for ML & DL

Algebra, Probability, Statistics, Calculus, Matrices

Education

M.sc Mathematics

BCU , Bangalore
2020 - 2022

6.5 sgpa

B.sc Mathematics

Gulbarga University
2017- 2020

72%

12th Board

Crescent P.U College.
2015 - 2017

78%

Language

English
French
Hindi
Urdu
Kannada

Interest

Cricket
Photography
Cooking

Frameworks & Libraries

Tensorflow, Keras, Pytorch , Transformers , Simple Transformers , Pandas, NumPy , SpaCy , NLTK , Matplotlib , Seaborn, Langchain, Huggingface , OpenAI, LLama, Git , GitHub

ML & DL Algorithms

Random Forest	ANN , CNN ,RNNN
KNN	Vgg , ResNet
Decision Tree	AlexNet, Lenet
K-Means	InceptionNet ,
DB-Scan	EfficientNet
	SSD ,Yolo ,R-CNN, Mask-RCNN, MTCNN

Tools

LabelStudio , Labelimg , Lableme,
Ubuntu , Pycharm IDE, Anaconda IDE,
Postman