

## Self-Introduction

My name is Keshava Sharma, and I am a Data Scientist with 4.4 years of relevant experience. My expertise lies in Generative AI, Computer Vision, Machine Learning, and Natural Language Processing (NLP). Over the years, I have worked on various models to solve business problems using a range of algorithms and technologies.

In Generative AI, I have hands-on experience in building chatbot to address the business needs of the client and provide customized solutions.

In the field of Computer Vision, I worked on developing a YOLO model for wound detection at AIIMS, New Delhi.

In Machine Learning, I am proficient in using classification and regression algorithms like Logistic and Linear Regression, Decision Trees, KNN, and ensemble techniques- boosting and Bagging.

I have experience in handling end-to-end projects starting from data collection, preprocessing, analysis, model building, validation, fine tuning using hyperparameters, model deployment, and presenting results to higher authorities.

I have experience in handling projects as per the stakeholder requirements, both as an individual and a team player, for completion of the project within the given timeline.

## GenAI Project

### Problem Statement-

The client was an online retail company. They wanted to make a chatbot which their customers can use to gain specific information about their products. This will lead to increase their customer satisfaction and operational efficiency. This chatbot will decrease their annual operation cost and reduction in calls in the call centre.

As the TEAM MEMBER, I was responsible for the overall Development, model optimization, and continuous improvement of the chatbot.

The core of the chatbot was powered by state-of-the-art generative AI models like GPT-3.5 that I fine-tuned extensively for natural language processing.

### Steps

After setting up the virtual environment and importing the required modules, we-

1. We used beautiful soup 4 to extract the data from web and pypdf2 for pdf.
2. We pre-processed the data using Recursive Character Text Splitter with a chunk size of 1000 tokens and 200 characters as chunk overlap.
3. We generated embeddings using OpenAI Embeddings.
4. Stored the embeddings in FIASS.
5. We built the RAG Pipeline using Lang chain Retrieval QA.
6. Set up chatbot to feed the query and retrieve relevant answer.

7. Fine-tuned the chatbot using domain-specific examples with prompts.
8. We deployed the model on AWS EC2 instance.

## Yolov5 Project

### Problem statement-

The task was to develop a wound detection model to identify different types of wounds. This model will assist healthcare professionals in tracking the type of wound, wound healing and incidences of infections.

### Dataset size-

### Steps-

1. We collected the data in the form of images from hospital records.
2. We annotated each instance with pixel-level masks, bounding box coordinates and assigned them class labels using labelme. A healthcare professional assisted in the annotation process.
3. We resized the images (640,640) and normalized the images to ensure consistency in the input data.
4. We applied data augmentation techniques like random cropping, flipping, rotation, colour and brightness adjustments to increase the diversity of the data.
5. We split the data into training, validation and test set (using train-test-split).
6. We trained the YOLOv5 model on the training set.
7. We evaluated the model's performance on validation set using metrics- IoU (Intersection over union) and mAP (mean average precision).
8. We experimented with hyperparameters of the model using learning rate, batch size and anchor-box size to optimize the model's performance.
9. The final testing after fine-tuning the hyperparameters was done on the test set.
10. Deployed the model on AWS cloud.

## Sentiment Analysis Project

This project involved a thorough examination of customer reviews to extract valuable insights regarding product preferences and opportunities for enhancement. By applying sentiment analysis techniques, the project aimed to measure customer satisfaction levels and contribute constructive feedback to the product development process.

Our client wanted to conduct sentiment analysis of retail product and classifying the reviews into positive and negative sentiment from that they want to measure customer satisfaction levels and contribute constructive feedback to the product development process

**Steps-**

1. The data was handed over by client in csv format.
2. I pre-processed the data by lowercasing, removing HTML tags, URLs, chat word treatment, removing punctuation and stop word removal.
3. Then I did lemmatization to bring the words in their lemma form.
4. I then used the TF-IDF vectorizer to convert words into vectors.
5. Then, we various algorithms for modelling.
6. We achieved highest accuracy using Logistic Regression.
7. Then we deployed the model on Streamlit.