

# Sankha Subhra Mukherjee

<https://www.linkedin.com/in/sankha-subhra-mukherjee-3358b811b/>  
<https://github.com/SANKHA1> | [SANKHA1.github.io](https://SANKHA1.github.io)

sankhasubhramukherjee@gmail.com  
+91 8910527124

## SKILLS

---

**Programming Languages:** – Python, R, SAS  
**Deep Learning frameworks:** – PyTorch, LangChain  
**Programming libraries:** – Tensorflow, OpenCV, Scikit, Seaborn, Numpy, Pandas, Matplotlib, PySpark  
**Configuration management:** – Git, SVN, BitBucket  
**Visualization Tools:** – Tableau, QlikSense, PowerBI  
**Databases:** – SQL server, MongoDB, MySQL, PostgreSQL  
**Cloud Platforms:** – AWS, Azure, GCP  
**Data Engineering Tools:** – AWS Glue, Airflow, Apache Beam, BigTable  
**API Frameworks:** – Flask, FastAPI

## WORK EXPERIENCE

---

### Chryselys | Business Analyst | Chennai

May 2022 – Present

- Worked in the **Healthcare** domain and created a **Machine Learning** model to predict the number of patients who are going to switch their drugs in the next 3 months using historical drug usage data and demographic data with **CatBoost** and **XGBoost**, model with 92% accuracy. Performed **Exploratory Data Analysis, Data preprocessing, Model building** and **Model Validation** for the same. Published the final report in **PowerBI** dashboard.
- Worked in the area of **Coversational AI**. Developed a chatbot using **LLAMA 2 7B** model with **0.9 BLEU score**.
  - i) Employed the **Sentence Transformers** model of **HuggingFace** to generate embeddings from textual chunks within documents.
  - ii) Established a **Vector Database**, organized into multiple collections for various studies, and storing all embeddings within using **ChromaDB**. The **LLAMA-Index** facilitated efficient querying and comparison of multiple documents.
  - iii) Integrated ChromaDB with **Langchain** and queried the vector database and got results using **Similarity Search**.
  - iv) Performed **Instruction finetuning** over the **LLAMA 2 7B** model. Used **RAG** based approach over the finetuned **LLAMA 2 7B** model and passed the results through it to get the outputs.
  - v) Deployed the model using **Amazon Sagemaker**. Created a **Custom API** to create the frontend of the chatbot.
- Worked on an **NLP project** for **Abstractive Text Summarization** of interview videos:-
  - i) Extracted audio from videos using **Moviepy** module and used **Google Speech Recognition** API to transcribe the audio files.
  - ii) Used the **Reinforcement Learning model GPT-3** and **BERT** to summarize the text.
- Created **REST APIs** using **Flask**. Automated **QlikSense** and **MS Excel** reports using **Python**.
- Developed a **Web Scrapping pipeline** to automatically log into customer website using **BeautifulSoup** and find a specific information from images using **Pytesseract**.
- Created a job using **Amazon CloudWatch** which will run a python script in **Amazon Sagemaker** once every week on a specified time and create report and send them to stakeholders.
- Performed **ETL Loading** using **AWS Glue**
- Worked on **SQL** development in **Amazon Athena** and **Snowflake**.

### Finarb Consulting | Machine Learning Engineer | Kolkata

Nov 2021 – March 2022

- Worked on an image classification project where I classified broken tablets from whole tablets in a pharmacy factory. Used **Tensorflow** in distributed training for parallel processing of **SSD Resnet** and **Faster RCNN** models increasing processing speed by **40%**.
- Created a Machine Learning model to predict if a machine in pharma factory is going to fail within next 7 days using historical sensor information with **90%** accuracy.
- Automated error detection from an ML model using **python**.

### Feynn Labs | Machine Learning Intern | Remote

August 2021 – October 2021

- Worked on **Market Segmentation** and **Customer Characterisation** using **K-Means clustering** algorithm for EV start ups.

### Tata Consultancy Service | Assistant System Engineer | Chennai

Nov 2017 – July 2019

- Worked in the **Retail** domain, created a **Machine Learning model** to do **Market Basket Analysis** for our client:-
  - i) Found out hidden association between products for better cross-selling and upselling.
  - ii) Performed **Customer Segmentation** for targeted marketing and anticipate customer behavior.
  - iii) Built a **Machine Learning** model to predict which previously purchased product will be in user's next order.
- Created a **Machine Learning** model to forecast weather and disaster in a specific place.
- Developed **Market Segmentation** algorithm for a hotel chain.
- Worked on **file automation** with **Python**.
- Worked on installation of **Docker** using **Docker toolbox**. Created custom **Docker container images**, **tagging** and **pushing** the images.

## EDUCATION

---

Master of Technology in Optical Engineering | Indian Institute of Space Science and Technology, Trivandrum July 2019 – June 2021

B.Tech in Electronics and Communication Engineering | Maulana Abul Kalam Azad University of Technology July 2013 – July 2017

## ACHIEVEMENTS

---

5 Star in Python and SQL in Hackerrank ([Link](#))

4 Star in Leetcode ([Link](#))

## PROJECTS

---

### Object Detection with YoloV5 and Deepsort and Tensorflow FER ([Source Code](#))

- Here I detected and counted the number of people with **YoloV5-deepsort** and their facial emotion using **Tensorflow FER** module in a video of a busy street.

### Personalized Chatbot using PyTorch ([Source Code](#))

- Here I tried to create a personalized chatbot using **PyTorch**.

### Data warehousing using AWS Redshift ([Source Code](#))

- Here I tried to create an **ETL pipeline** to populate a database in **AWS Redshift**.

## CERTIFICATION

---

Data Engineering Big Data and Machine Learning on GCP ([Link](#))

AWS Fundamentals ([Link](#))