

SIDDHESH LIHE

DATA SCIENTIST / AI-ML ENGINEER

PHONE :- +91 7208249266

E-MAIL :- sidlihe88@gmail.com

LinkedIn Profile :- <https://www.linkedin.com/in/siddhesh-lihe-711104212/>

GitHub Link :- <https://github.com/sidlihe>

Mumbai, India

PROFILE SUMMARY

Experienced data scientist skilled in stock market analysis, OCR integration (Gujarati, Hindi, Sanskrit), and daily data retrieval via APIs. Expertise in Python, ML, deep learning, NLP, Gen AI, CICD, AWS, and Agents. Proven track record in custom detection and recognition using ensembling of classification and image-to-text methods. Proficient in deploying models with Flask, Streamlit, and visualizing data.

SKILLS

Python :- Pandas, NumPy, Matplotlib, Seaborn, Mathematics , FastAPI

Machine Learning :- Supervised Learning Algorithms, Un-supervised Learning Algorithms , MLFLOW

Deep Learning :- TensorFlow, Keras, Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN)

Deployment :- Flask, Streamlit ,Dockers , AWS, Git-CICD

Natural Language Processing:-

Text Pre-processing – Extracting text, Tokenisation, Stemming, Lemmatization, Text Cleaning using NLTK Packages, Information Extraction and Preprocessing

Word Embeddings – TF-IDF and Word2Vec , FAISS,

Data Visualization Tool :- Matplotlib, Seaborn Tableau, Power BI, My SQL

EDA :- Data Cleaning, Data Balancing, Data Wrangling, Data Visualisation, Data Augmentation

Gen AI:- Chat-Completion, Langchain, Chain & Prompt Engi., LLM, RAG, Agentic-AI

WORK EXPERIENCE

Nutriiya – (AI-ML Engineer.)

JULY 2025-CURRENT

- Developed a custom object detection pipeline for Indian cuisine using ResNet architecture, CLIP embeddings, and ensemble learning
- Built a recommendation system tailored to user preferences and regional food trends, enhancing personalized food suggestions.
- Designed and deployed a chatbot capable of handling FAQs and dynamic queries using Retrieval-Augmented Generation (RAG) with agentic capabilities for contextual reasoning.

- Integrated OCR with support for Gujarati, Hindi, and Sanskrit languages. Developed and implemented custom image processing techniques, including layout segmentation, text detection, text classification, and text recognition. Beat Tesseract's accuracy by optimizing the OCR pipeline for better performance and reliability. Enabled efficient extraction of textual information from documents and images with enhanced precision.
- Test APIs using Postman to retrieve daily commodity market data. Automated the daily capture and storage of market data for 3 million vegetables, fruits, and commodities, enabling valuable insights and analysis.
- Utilize Selenium and Beautiful Soup for web scraping financial data & automate it. Analysed and interpreted trends in the stock market and back-tested various strategies. Generate actionable insights for financial decision-making and building algorithms.

TCS – (NLP Intern.)**SEP 2022-DEC 2022**

- Analysed social media post sentiments using Sentiment Intensity Analyzer and TextBlob.
- Developed a sentiment analysis tool to track public opinion on specific topics.

Barrier Break – (Trainee Accessibility Tester)**APR 2022-MAY 2022**

- To test a webpage is accessible for a visually impaired people to access a webpage through Andy tool, CCA, HTML, CSS

PROJECTS

1. Business Card Reader (NER)
2. Snake Game using Python
3. Emojify
4. Multi-class image classification model using CNN architectures
5. MCQ generator using Gen. AI
6. PDF-Summarizer (LangGraph)

EDUCATION

MSC-AI B.K.BIRLA COLLAGE OF ARTS, COMMARCE AND SCIENCE AUTONOUMS, KALYAN(W)
JUN 2021 – MAY 2023 (CGPA 9.20)

BSC-IT B.K.BIRLA COLLAGE OF ARTS, COMMARCE AND SCIENCE AUTONOUMS, KALYAN(W)
JUN 2018 – MAY 2021 (CGPA 7.40)

PUBLICATIONS, HONOURS AND AWARDS

- Published article on Sensor Data Acquisition on the portal ([455757.pdf](#))
- Complete Machine Learning, Deep Learning, Reinforcement Learning course with TCS
- Successfully completed FULL STACK SOFTWARE ENGG. Course Through NIIT.