# Susanket Sarkar

Data Scientist

sarkar.susanket@gmail.com • +91-8777856661 • LinkedIN • Github • Kaggle

## Summary

Data Scientist with 3+ years of combined internship and full-time experience in building and deploying real-world AI system across image, audio, geospatial and text data, driving efficiency, accuracy, and tangible product impact at scale.

## Skills

**Python, R, SQL,** PyTorch, TensorFlow, Keras, **Docker**, Git, PowerBI, **AWS (S3, SageMaker), Azure, LLM, MLOps**

## Work Experience

**Data Scientist**, *AEREO* , Bengaluru, India — May 2024 - Present

- Built a **3D segmentation model** using drone data to detect mining features like haul roads, stockpiles, replacing manual mapping
- Developed models on **AWS** and **Azure AI** using **Dockerized batch jobs**, allowing on-demand scaling across enterprise workloads
- Designed and maintained **MLOps** pipelines, from training to deployment and monitoring, ensuring reliability, speed in production
- Created a full **Python**-based repository for all pre and post-processing steps, including data cleaning, image slicing, from scratch
- Used **SQL** to build dashboards and metrics so business teams could track model outputs, performance, and validate predictions
- Integrated **OpenAI's GPT-3** into a **natural language dashboard**, allowing users to ask questions and get these translated into **SQL** and spatial queries, enabling intuitive, conversational access to geospatial insights.
- Enabled spatial query support by connecting **GPT-3** to **PostGIS database**, allowing retrieval and analysis of mining infra metrics
- Developed rural infra models detecting rural roads, lakes, trees, wells, poles, and buildings using orthoimages with **96%** accuracy
- Automated pipelines via ML, cutting manual digitisation time by **93%**, boosting accuracy by **45%**, and cutting **$10.8K** in ops cost
- Bridged technical gap by crafting metrics and dashboards, enabling **stakeholders** to interpret model insights and validate outputs

**Data Science Intern**, *AEREO*, Bengaluru, India — May 2023 – July 2023

- Directed a project for micro-object detection in images, solving business problems through aerial analysis using machine learning
- Managed a vast **3TB** photogrammetry dataset sourced from **150** distinct vendors, streamlining data organization and storage
- Developed a MXNet model with **92%** precision and drove automation, saving **Rs. 10L** in annual expenses and **40** weekly hrs
- Integrated advanced features into Aereo's B2B SaaS Analytics; uplifting the Platform Adoption Rate by **27%**, increasing growth

**Data Science Intern**, Fitbuddy, Bengaluru, India — Sept 2022 - Nov 2022

- Leveraged Python's OpenCV & MediaPipe libraries to form logic of a Body Pose Detection Model enabling body pose detection
- Formed relational matrices by considering landmark relationships and trained the system on Naive Bayes, SVM, Random Forests
- Refined and optimized codebase to detect exercises accurately, adding functionalities for yoga poses, obtaining an **87%** F1 Score
- Integrated the feature to production, resulting in a **20%** increase in application downloads due to superior pose detection UX

**Data Science Intern,** Vitt.AI — January 2023 – March 2023

- Automated a system that transcribed and translated **18 hrs** of speech data from 10 languages from customer service conversation
- Utilized MFCC (Mel Frequency Cepstral Coefficients) for audio feature extraction, preparing data for transcription and translation
- Finetuned Whisper model on audio data, achieving a Word Error Rate (WER) of **<19**, translating conversations to English.
- The pipeline witnessed an increase in efficiency by **15%**, automating tasks and saving **27** hours in language processing endeavors.

**Data Science Intern**, Kincare — July 2022 – September 2022

- Engineered and deployed a data pipeline for an in-app chatbot, facilitating users to upload and query their PDF medical reports
- Developed dashboard and visualizations using Power BI, showing **30**+ vital KPIs and providing insights to business stakeholders
- Automated **10**+ tasks leading to **52**% reduction in analysis time, streamlining the process to solve the business problems efficiently

## Projects and Research Work

**Optimising Latent Representations in VAE** (Supervisor: Prof. S. P. Pal, IIT Kharagpur) — Jan. 2024 – Apr. 2024

Built a VAE-LSTM framework on battery datasets to test denoised latent representations; improved prediction accuracy by 30%, achieving an $R^2$ of 0.93, proving constrained latent spaces can enhance sequence learning and extract robust temporal patterns.

**Explainable Deep Learning for Heart Sound Analysis** (Supervisor: Prof. S. P. Pal, IIT Kharagpur) Aug. 2023 – Dec. 2023

Designed an explainable deep learning pipeline for cardiac disease detection using a fusion of auditory features and fine-tuned models; achieved 99.1% accuracy, ensuring transparency through interpretable visualizations (LIME, Shapely) and decision paths.

**Real Time Forest Fire Detection** (Supervisor: Prof V. Rachella, IIT Kharagpur) — May 2022 - August 2022

Developed a forest fire detection system using a custom ResNet50 pipeline trained on 25K images; achieved 92% accuracy and 87% F1 score, integrated alert system via SMTP, and successfully flagged 3 live fires, enabling early intervention and hazard control.

## Competitions

**Ranked 7/5255 at Trilytics '23 (IIM-C)**: Built an **87%**-accurate risk prediction system & dashboard to forecast mining accidents and business impact.

**Won Cisco Convolve '23 (1/1500):** Built an inventory demand predictor and management system using SARIMAX & Prophet, reducing SMAPE by **23%** and forecasting a year ahead.

## Education

**Bachelor of Technology**, *Indian Institute of Technology, Kharagpur* — Nov 2020 – April 2024

*Major: Aerospace Engineering (Specialization in Artificial Intelligence and Machine Learning)* — GPA – 8.06/10

*Relevant Coursework*:

- **Statistics for AI/ML**: Sampling, Mixture Models, Hypothesis Test, Reinforcement Learning, Causal Inference
- **Machine Learning**: Bayesian Theory, Decision Tree, Ensembles, Support Vector, Dimensionality Reduction, Perceptron, CNN, RNN