Susanket Sarkar

Data Scientist
sarkar.susanket@gmail.com | +91-8777856661 | Kaggle | LinkedIN | Github

Skills

Python, SQL, **R, C/C++,** Numpy, Pandas, AWS (EC2, S3, Sagemaker), GIT, Docker, Pytorch, TensorFlow, Keras, PowerBI

Work Experience

AEREO Bengaluru, India
Data Scientist **May 2024 - Present**
AEREO leverages geospatial aerial imagery to develop intelligent solutions for mining sites and rural areas, focusing on feature detection and optimised planning from a single image source.

- Developed a mining road detection algorithm integrating Dijkstra-based pathfinding and clustering to identify safe, efficient haul roads for dumper trucks based on origin, destination, and terrain maps, optimising route efficiency.
- Achieved precise road-edge detection with **89% IoU** against ground-truth data, improving **accuracy** by **39%** and reducing path planning time by **93%**
- Enhanced route optimisation for mining vehicles by factoring in metrics like travel time and slope analysis, contributing to a **5%** increase in project adoption
- Designed scalable workflows integrating statistical learning methods for aerial geospatial imagery. Improved mining route
- detection using advanced graph theory and optimisation techniques
- Deployed machine learning techniques to datasets, reducing digitisation time by **99.5%,** Improved digitisation accuracy by **45%** compared to existing AI models and cut processing time by **25%**, reducing costs by **$10,800**
- Boosted client adoption by **14%** due to the combined improvements in both mining and rural digitisation workflows

Data Science Intern **May 2023 – Jul. 2023**

- Worked with over **30GB** of data from **150+** geographical sites, efficiently queried data stored in **S3**, enhancing the scalability and speed of data processing tasks
- Developed **multiple deep learning models** tailored to various tasks, **addressing data scarcity** challenges, and achieved an average precision of **92%** and recall of **89%** across all objectives
- Productionised features, resulting in cost savings of **$10K** annually, and reducing weekly workload by over **40 hours.**

Fitbuddy New Delhi, India
Data Science Intern **Aug. 2022 – Nov. 2022**
A fitness technology startup focused on leveraging AI to provide real-time exercise feedback and personalised fitness insights through innovative applications.

- Developed AI application for autonomous body pose detection and real-time corrective feedback using **Python, OpenCV,** and **MediaPipe**; integrated machine learning algorithms (**SVM, decision tree, LSTM**) to predict incorrect poses
- Expanded functionality to support **18+** yoga poses, achieving an **87% F1 score**; deployed the scalable model to production, driving a **20%** increase in app downloads
- Optimised user engagement, resulting in a **15%** increase in interaction rates and **20%** increase in application downloads

Kincare Bengaluru, India
Data Science Intern **Nov. 2022 - Dec. 2022**
Kincare specialises in creating user-friendly, AI-powered applications for elderly care, enabling medical report analysis and other features.

- Developed a data pipeline for an in-app chatbot for PDF medical report analysis and question answering for elderly users.
- Designed dashboards in **Power BI**, visualising **30+** KPIs and presenting insights to **40+** stakeholders, driving data-informed decisions.
- Automated **10+** manual processes, reducing analysis time by **52%** and increasing app engagement using UI optimisations.

Research Work

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

- Investigated the effectiveness of constrained encoder representations in Variational Auto-encoders (VAE) to extract non-trivial, high-value features, addressing limitations of reconstruction-based objectives prone to trivial solutions
- Developed a VAE-LSTM pipeline to evaluate the interplay between latent space denoising and sequential learning dynamics, leveraging time-series battery datasets as a validation framework
- Achieved a **30%** enhancement in predictive accuracy, attaining an R^2 score of **0.93**, substantiating the role of representation regularisation in improving downstream time-series forecasting tasks

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

- Explored explainable deep learning architectures for heart sound classification, focusing on model transparency.
- Developed a custom Fusion-based Disease Classification (FDC) framework, integrating transfer learning from multiple pre-trained models and leveraging auditory features such as Spectrogram, MFCC, and chroma-gram.
- Achieved **99.1%** accuracy, validating the model's robustness and generalisation capability in diverse heart sound datasets.

Education

B. Tech, *Indian Institute of Technology (IIT), Kharagpur* (GPA – 8.02/10)
Relevant Coursework:

- **Statistics for AI/ML**: Sampling, Mixture Models, Hypothesis Test, Information Retrieval , Reinforcement Learning, Causal Inference
- **Machine Learning**: Bayesian Theory, Decision Tree, Ensembles, Support Vector, Dimensionality Reduction, Perceptron, CNN, RNN
- **Advanced Learning Paradigm**: Semi supervised learning, Few-shot learning, Multi-task Learning, Meta Learning, Transfer Learning

Extra-Curricular Activities

- **Volunteering** at **National Social Services** (NSS) - Led education camps in villages and coordinated aid distribution for community development.
- **Trilytics 2023 Hackathon** at **IIM Calcutta** (Ranked **4**th out of 1200) – Developed a risk prediction framework (87% accuracy) and customer-centric dashboards for accident forecasting and threat analysis.