Arkadip Maitra

+919775940390 arkadipmaitra@gmail.com arkadip-maitra.github.io linkedin.com/in/arkadip-maitra/

Education

Ramakrishna Mission Vivekananda Educational and Research Institute **Sep 2021 – June 2023**
Master of Science in Computer Science *8.70 CGPA*

Ramakrishna Mission Vivekananda Centenary College **Aug 2018 – June 2021**
Bachelor of Science in Computer Science *9.09 CGPA*

Work Experience

PrediQt Business Solutions Pvt. Ltd **Kolkata, India**
Senior AI/ML Engineer *Oct 2024 - Present*

- Fine-tuning large language models using collected and open source data for tasks like product categorization, description generation from product title and images, filtering user query, filtering products returned from vector store and generating condensed query based on chat history and by comparing results with models from OpenAi and Anthropic

Associate Data Scientist *July 2023 - Sept 2024*

- A virtual shopping assistant. This RAG based solution can recommend products and answer questions from ingested data for any user query. Built using multiple LLM calls, Weaviate as the vectorstore, MongoDB for storing chat history, FastApi for backend and managed with kubectl, running on a Kubernetes cluster. Link
- A virtual real-estate agent. This RAG based solution can describe and recommend properties based on user query from all over the USA. It filters out properties based on location given in natural language input. Built using multiple LLM calls, Weaviate as vectorstore, MongoDB for storing chat history, FastApi for backend and managed with kubectl, running on a Kubernetes cluster. Link
- Predicting winner of each Euro 2024 match. Various ML models like Random Forest, SVM, XGBoost was trained on previous match data. The best model Random Forest predicted **39** out of the **51** Euro 2024 matches correctly. This application enabled site visitors to cast their votes for the winning team and limited each ip address to one vote per match. This application used the trained ML model for predicting winning chances, MongoDB for storing user vote data, FastApi for backend and was hosted on Railway.

Data Analyst Intern *Feb 2023 - June 2023*

- Created a imperceptible document watermarking solution. A fine-tune of Stegastamp was done using open source document image datasets. It reached watermark decryption accuracy of **80%** on 30 well lit watermarked documents. Used PyTorch for model development, Flask for backend and locally hosted for testing.
- Predicting Jira support ticket volume raised on any particular date. Various models like Prophet, Neural Prophet, LSTM and Transformer were tested. An LSTM variant attained lowest MSE of **0.27**. This solution was developed using PyTorch, Flask for backend and is hosted on local server for internal use.

Kyrion Technologies **New Delhi, India**
Android Developer Intern *June 2019 – July 2019*

- Assisted in development of an android application feature that tracks user attendance at any particular location based on user phone location and a selfie. Application uses Google Geolocation API for location verification and Java for android code.

Research Experience

Sign Language Generation **Aug 2024 - Present**
Associate Researcher *Dept. of CS, RKMVERI*

- Supervisor: Prof. Soumitra Samanta
- Working on developing a sign language generation which creates sign video from input sentence.
- A body part based modification of vqvae is able to create almost exact replication of input skeleton motion with reconstruction loss below $\mathbf{7e^{-4}}$ for both Phoenix14T and How2Sign datasets.

Sign Language Recognition **July 2023 - June 2024**
Associate Researcher *Dept. of CS, RKMVERI*

- Supervisor: Prof. Soumitra Samanta
- Assisted in collecting and processing of 40,000 Indian sign language videos, for creating the FDMSE-ISL dataset.
- Developed a novel body part based graph attention network called Hierarchical Windowed Graph Attention Network.
- HWGAT attained state of the art results for keypoint based methods across five different datasets. Achieved **Top-1** accuracy of **97.79** for INCLUDE, **93.86** for FDMSE-ISL, **98.59** for LSA64, **95.80** for AUTSL and **48.49** for WLASL datasets. Code & Website

Self Supervised Writer Identification **July 2022 – Dec 2023**

Research Intern *CVPR Unit, ISI Kolkata*

- Supervisors: Prof. Umapada Pal, Prof. Saumik Bhattacharya and Siladittya Manna
- A decorrelation-based self supervised learning technique for writer identification from handwritten text image. Code
- Achieved **Top-1** Word level accuracy of **84.8** for IAM, **93.32** for CVL, **74.24** for Firemaker datasets and **Top-1** page level accuracy of **95.58** for IAM **96.87** for CVL, **98.40** Firemaker datasets.
- These results were on par with other supervised techniques and better than existing self supervised methods.

Earthquake Early Warning System **Oct 2020 – Nov 2021**

Student Researcher *Dept. of IT, University of Calcutta*

- Supervisors: Amlan Chakrabarti
- Extracted seismic features from time-series seismic data collected by sensors all over the Indo-Himalayan region.
- These parameters are then used to classify whether the seismic activity will require alarm which has threshold 5.5 magnitude.
- This early warning system uses these parameters obtained from the fast arriving p-wave to raise alarm if it classifies the slower s-wave to be destructive.
- Depending on the distance from epicenter, the speed difference of 1.5km/s to 3km/s gives enough time for people to seek shelter when advanced alarm is raised.

Publications and Preprints

Suvajit Patra, **Arkadip Maitra**, Megha Tiwari, K. Kumaran, Swathy Prabhu, Swami Punyeshwarananda, Soumitra Samanta Hierarchical Windowed Graph Attention Network and a Large Scale Dataset for Isolated Indian Sign Language Recognition. *Multimedia Tools and Applications* 2024 (Under Review)

Arkadip Maitra, Shree Mitra, Siladitya Manna, Saumik Bhattacharya, Umapada Pal Decorrelation-based Self-Supervised Visual Representation Learning for Writer Identification. *Transactions on Asian and Low-Resource Language Information Processing* 2024 (Accepted. Yet to be published)

Samik Basu, Sayan Tripathi, Soumen Halder, **Arkadip Maitra**, Pritha Banerjee, Amlan Chakrabarti Enhancing Earthquake Preparedness in the Himalayan Region: A Machine Learning Approach using EEW System Parameters. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering* 2024 (Under Review)

Samik Basu, **Arkadip Maitra**, Soumen Halder, Soumya Pandit, Soma Barman, Pritha Banerjee, Amlan Chakrabarti Machine Learning Based Earthquake Early Warning (EEW) System: A Case Study of Himalayan Region. *International Conference on Data Management, Analytics & Innovation* 2022

Anubrta Das, Soumen Halder, **Arkadip Maitra**, Shree Mitra and Raj Sen A Method for Artifacts Removed from MRI of Brain. *European Journal of Pharmaceutical and Medical Research* 2019

Miscellaneous

Delivered a lecture series on Git, Docker, and API hosting for deploying trained models in the *Deep Learning & Natural Language Processing* (DA345) course.

Core team member of the RKMVERI tech fest Perceptron 2023. Created and hosted coding competition.

Second runner up of the Envision 2019 coding competition hosted by RKMRC, Narendrapur.

Certificates and Courses

Generative Adversarial Networks (GANS) Specialization from Coursera.

Deep Learning Specialization from Coursera.

Introduction to Programming in Java from Microsoft

Natural Language Processing by Prof. Soumitra Samanta

Approximation and Online Algorithms by Prof. Subir Kumar Ghosh

Algorithms for Data Science by Prof. Anil Maheshwari

References

Prof. Soumitra Samanta Email

Swami Punyeshwarananda Email

Abhishek Nandy Email

Siladittya Manna Email