Soumyabroto Banerjee

Manager AI

📱 (+91) 94775 55030

🌐 soumyabrotobanerjee.github.io
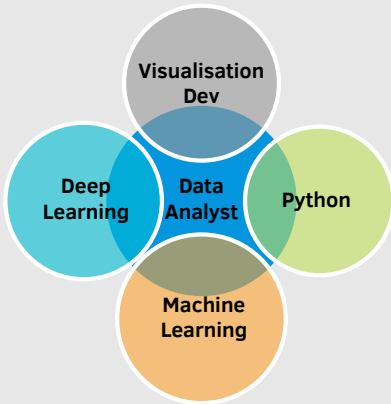
✉ soumyabroto.banerjee@gmail

in /in/soumyabroto-banerjee

⬤ SOUMYABROTOBANERJEE

Technical Skills

Overview

Programming

0 *LOC* ⟶ 5000 *LOC*

- Python

- Computer Vision • NLP

- Qlik Sense • Tableau • Power BI

Education

B. Tech, Electronics and Communication (GPA: 9.12)
IRPEL, University of Calcutta
2016 - 2020 | Kolkata, India
Achievements:

- **TATA Milennium Scholar - WBJEE Rank 1084**

- **STEX 2018, Runner's up (Tech Fest, IEEE)**

- IEEE EDS Member - Organised Seminars and Events

Experience

Jan 2022 - Present

Manager AI, Tata Steel Ltd   Tata Group Transfer (Not Job switch)   TATA STEEL

- Responsible for bringing **Digital Transformation** through the use of **AI** and Advanced Cognitive Algorithms. Automating and Articulating Processes at simplistic efficiency to add to the Steel Value Chain from **Blast Furnace Twins** to **Hyper Personalised D2C Websites**.

  Some Projects:
- **Process Brain** - **Autonomous Blast Furnace Operation**s using Latent Space modelling of Blast Furnace Regimes.
- **Hyper Personalization of B2C Zone** - Aashiyana (aashiyana.tatasteel.com) through Collaborative Filtering
- **Competitor Profiling** - Track, analyze and monitor **Competition profiles through Web Scraping, Social Media Listening, etc**

Aug 2020 - Jan 2022

Systems Engineer, TCS Digital   TCS

- **Qlik Sense Visualization Lead** on the client billable side. Business Requirement identification. Developing **Javascript** based **Extensions** and Themes for a better user experience. Using **Agile** methodologies and Effective **DevOps** Model for best and optimized Development and Delivery.
- **Augmented Chat Bot Developer** - **NLP** Based Chatbot built from **RASA** in **Python and JS**, it can connect to various platforms like Qlik, Tableau, Power BI and pull your visualizations. It can analyse and develop stories for your data on the fly through its own algorithms. **Microservices** based architecture which is completely **modularized** and built on **Open Source Softwares**. It can connect to DWHs and DBs to pull data and analyses.
- **CoE, Data Science** - Active Member and Data Scientist at Centre of Excellence - Data Science. Worked on multiple POCs to identify and mitigate Client Pain Areas through the magic and language of data. POC include - Covid MRI Segmentation, Future News Prediction
- **Big 5 Video Analyzer** - First Impression (ChaLearn et. al.) to analyze a Video **undersampled by 0.30** to produce the overall Big 5 Traits of a Subject in Question. Trained and Tested for an accuracy of **0.98 on CNN Based Network**.
- **NLP Based MBTI and Big 5** - Developed 14 NLP Models - based on **Naive Bayes, LSTM Based, RF and Decision Tree based**. MBTI classes were trained in an **OVO manner**. The MBTI based **LSTM Network outperformed** other models and it incorporated all the classes as well. For Big 5, we trained a simple Naive Bayes and also a **sequential model to add to the Video Analyzer**. Checked for correlation between the two domains - MBTI and Big 5. Overall Accuracy above **0.88**.
- **QnA Chatbot** - Chatbot based on **Image Captioning, CLIP, GPT 2 and Elastic Search**, it can understand a Question a student wants to ask just **from an image and/ or extract texts** around that image. Works on a feedback loop to confirm the context of questions and uses a **robust fuzzy based Elastic Search to search contextual answers** from a connected database.
- **Mask Based Face Recognition** - This was the major challenge where the person doesn't need to open his mask for face detection to work. It Works by Latent based Facial Feature Extraction. Acc: 91.2

| Jan 2020 - Apr 2020 | **Deep Learning Personal** | |

- **Implementation of StyleGan in TF2.0+, architecture of Nvidia.**
- **Latent Space Control** Control of Latent Space to control the generation of Images.
- **PreTrained VGG** To Encode StyleGan Generated Images to Latent Space
- **Generating Controlled Images** - Synthesis and Discrimminator Network to control generation of Images close to VGG output.
- **Blending like FaceApp** - Blend Images with weights to get Facial Blendings on Embedding level.

| June - 2019 Aug - 2019 | **Machine Learning Intern** | Tata Steel Ltd |

- **Prediction of high delay cash to invoice flow**. Identification of High Delay Zone Customers.
- **Base Model** was adopted as **SARIMAX** to build on top of it.
- Built end to end **Machine Learning** Pipeline with **Bagging and Boosting** for Final prediction.
- **Imbalance** was tackled using **sampling** techniques
- **LSTM** based model used to process the time dependency.

Research

| 2019 - 2020 | **B.tech Candidate, Research Student** | University of Calcutta |

Thesis: CNN based Multimodal Brain Tumour Detection
- Using **SWT** and **PCA** to merge and blend MRI and SPECT Images to extract maximum information from both worlds.
- **CNN for automatic feature extraction** and using visualization of CNN firing patterns to determine the features learnt on classification.
- Received Great Appreciation and **1st Best Paper Award in AMPHE, 2020**.
- **Proceeds published in Springer Nature Journal, 2021. (click)**
- **Tools**: Python, numpy, Tensorflow, scikit-learn, pandas

| 2019 - 2020 | **Research Intern** | Jadvapur University |

Thesis: Novel Clustering Algorithm
- Converting of **weak hard-cluster based algorithms** to **fuzzy algorithms** using a probabilistic model
- **Moving of cluster points in a defined cluster space** for **outlier** removal and better clustering **shape index**
- Proceeds yet to be published
- **Tools**: Python, numpy, scikit-learn, pandas

About

Why Me ? I stick to problems till they get solved and I am never disappointed if they don't because I learn a lot if they don't as I know the dark lanes the next time I face a similar challenge.

Data and its Science has always been like Chemistry to me. It is mysterious, yet so fascinating that you cant but just enjoy it.

I look for challenges in my everyday work and newer challenges keep me motivated. I am an obsessive person who seeks the best out of everything, constantly learning in the process. If I finish something, I deliver it, but in my free time go back to it and check for better, viable options. I love Multi Tasking and working on multiple tasks feels like I can utilize most part of my skill set.

Data is something that motivates me because data gives one the power to actually dissociate lies from the truth. It gives a sense of oneness to solve problems that are otherwise not so easy to solve.

I have been an active learner and an active follower of everything data. I love to work on data and let data do more of the work through Python Codes.

I am an extremely positive person with a never ending thirst for knowledge. I also do have excellent oratory skills, do let know if we are up for a debate.