Soumyabroto Banerjee

AI Engineer

(+91) 94775 55030

soumyabrotobanerjee.github.io

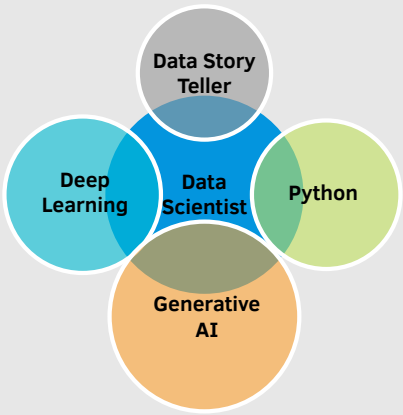
soumyabroto.banerjee@gmail

/in/soumyabroto-banerjee

SOUMYABROTOBANERJEE

Technical Skills

Overview



Programming

$0\ LOC \longrightarrow 5000\ LOC$

- Python
- Computer Vision • NLP • Gen AI
- Qlik Sense • Tableau • Power BI

Education

- **B. Tech, Electronics and Communication *(2016 - 2020)***
  IRPEL, University of Calcutta
  - **CGPA: 9.12**
  - **TATA Millennium Scholar -** WBJEE Rank 1084
  - **IEEE STEX 2018, Runner's up**
- **School: Don Bosco|** Kolkata, India
  - **ISC: 94.5% - ICSE: 91.8%**

Experience (3.5+ Years)

Jan 2022 - Present — **Manager AI, Tata Steel Ltd** — TATA STEEL

- Responsible for bringing **Digital Transformation** through the use of **AI**. **Critical** in **setting up the AI Practice** at TSL.
  Projects:
- **Enterprise GPT**
  - **A scalable, private and secure version of ChatGPT** developed for TSL.
  - **Ideated, Developed, Deployed & Maintain single-handedly**
  - **Generate insights from unstructured data like PDFs,** from **Structured data** like Plant health KPIs, and from **semi-structured** data like Safety observation and customer feedback.
  - Democratized insights to **Sr. Leadership and MD** with automated weekly mailers
  - Presented and published in Tata Sons as the ***"Flagship model for adoption and implementation of Gen AI"***
- **Process Brain - Autonomous Blast Furnace Operations**
  - Predicts the KPIs like Heat Loss, Hot metal temp, etc using Transformer architectures
  - The KPIs are pushed into an optimizer with constraints to predict operational set points for the most efficient operations.
- **Hyper Personalization of D2C Cust, Aashiyana**
  - Collaborative Filtering based nudges & reccos delivered a 15% increase in conversion
  - Working on a Transformer based approach to capture user segments.
  - Responsible for end-to-end delivery
- **Competitor Intelligence Platform**
  - **Competitor Data Repo:** Track, analyze and monitor **competitors** through Web Scraping, Social Media Listening, and News from paid articles.
  - **Using Gen AI,** delivered as a Chat Interface to generate insights contextual to Tata Steel.
  - Used extensively by **the Strategy planning Team**

Aug 2020 - Jan 2022 — **Systems Engineer, TCS Digital** — TCS

- **Qlik Sense Visualization Lead**
  - Business Requirement identification.
  - Developing **Javascript-based Extensions and Themes for a better user experience.**
- **Augmented Chat Bot Developer -**
  - **NLP** Based Chatbot built from **RASA** in **Python and JS**,
  - Connects to various platforms like Qlik, Tableau, Power BI
  - **Microservices** based architecture
  - (elected for Qtr Finals of Tata Innovista 21
- **CoE, Data Science**
  - Worked on multiple POCs to identify and mitigate Client Pain Areas
  - POCs include - Covid MRI Segmentation, Future News Prediction
- **Big 5 Video Analyzer**
  - CNN based on First Impression (ChaLearn et al.) to analyze a Video to produce the overall Big 5 Traits of a Subject in Question.
  - Trained and Tested for an accuracy of **0.88 on CNN Based Network**.
- **QnA Chatbot**
  - **Multimodal Chatbot** based on **Image Captioning, CLIP, GPT 2, and Elastic Search**.
  - **Can answer questions from both images** and text from an existing database indexed on ES.

Jan 2020 - Apr 2020 **Deep Learning Personal - Others** Personal Project (Click Here)

- **Implementation of StyleGan in TF2.0+, the 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 the generation of Images close to VGG output.
- **Blending like FaceApp** - Blend Images with weights to get Facial Blendings on the Embedding level.

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

- **Prediction of high delay cash to invoice flow**. Identification of High Delay Zone Customers.
- **The 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 (2+ Years)

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.

- **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 and Presentations.** Yeah, I love making presentations as well and have been quite fondly regarded as ***"PPT Man"***
- **Data and its Science** has always been like Chemistry to me. It is mysterious, yet so fascinating that you can't but just enjoy it.
- **I look for challenges**, a **multi-taske**r, obsessively compelled to make **better the best!**
- **Data 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.