Soumyabroto Banerjee

Data Analyst

Kolkata, West Bengal
9477555030
soumyabroto.banerjee@gmail.com
soumyabrotobanerjee.github.io

Summary

Data 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 skillset.

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.

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.

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

Work experience

Systems Engineer, TCS Digital

Aug, 2020 - Present

Tata Consultancy Services, Kolkata

Involved in multiple projects and POCs. Always Diligent to the task at hand and eager to learn more. Taking up newer challenges everyday to work on a wider and deeper panorama of Technologies.

- **Qlik Sense Visualization Lead** on the client billable side. Involves Identifying Best Fit Business **Visualization Requirements** and Developing them on Qlik Sense with a story to support the Business Questions. Involves 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 scratch, it can connect to various third party 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 Software. 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 - **Advanced Covid MRI Detection**, **Future News Prediction (current), etc**

Deep Learning Developer - Freelancer

May, 2020 - Sep, 2021

aQb Solutions

Working on multiple projects in multiple domains - including NLP and CV. Diligence Delivery of all requirements and modifications within timeline.

- **Big 5 Video Analyzer** - First Impression (ChaLearn et. al.) to analyze a Video undersampled by 30% to produce the overall Big 5 Traits of a Subject in Question. Trained and Tested for an accuracy of 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 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.

Deep Learning Developer - Freelancer   Apr 2020 - Aug 2020
Adben Industries

Development and Deployment of end to end Face Recognition and Mask Detection Framework when COVID induced No Touch Attendance System was a scarcity and a need at the same time.
- **FaceNet based Face Recognition** System to Register and track faces in real time.
- **Anti-Spoofing Engine** to detect whether Person in front of camera is a real person or an image.
- **Mask Detection Model** to detect masks and log into database if and when there are defaulters.
- **Mask Based Face Recognition** - This was the major challenge where the person doesn't need to open his mask for face detection to work. Works by F**acial Feature Extraction and Training**.
- **Cloud Based SaaS** - Coded from Scratch bypassing Streaming Software and deployed on GCP and AWS as a SaaS Model for Customers.

Received huge appreciation for timely delivery of Quality Software with all the requests/requirements fulfilled.

Student Researcher   May 2019 - Aug 2020
University of Calcutta

Worked on Multimodal Brain Tumor Detection using MRI Scans.
- Using **SWT** and **PCA** to merge and blend **MRI** and **SPECT** Images to extract maximum information from both worlds.
- Designed **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 to be published on Springer Nature Journal in July 2021.**

Machine Learning Intern   June, 2019 - Aug, 2019
Tata Steel, Kolkata

Duly Reporting and Deliverance of Projects and writing production level code for easy to deploy and debug.
- **Proposes a 5 way model with Bagging and Boosting** Classifier to predict the Sales growth, Prediction of Red Delay Customers and Accumulation for the Next Week.
- **Uses an on Oversampler** to tackle the problem of Imbalanced Dataset..
- A baseline of **SARIMAX** used.
- **LSTM** Based Network to understand the sequential data of the next feasible investment for the Supply Chain Accumulation
- **Future Pattern of Order Prediction** using Market-Basket-Analysis as a baseline for Tata Pravesh.

Education

Bachelor of Technology   Aug 2016 - Aug 2020
IRPEL, University of Calcutta

Overall GPA : 9.12
Recipient of : TATA Millenium Scholarship for WBJEE Rank of 1600
Member of IEEE EDS society - Helped Organize Seminars and Events
STEX, 2018 (Tech fest, IEEE) - Runner's Up

ISC & ICSE

Don Bosco, Park Circus

ISC, Science ; Marks : 94.5%

ICSE, Science ; Marks : 91.8%

Overall Skills

Deep Learning & Machine Learning

CNN, LSTM, Transformers, GANs,SVM, Bagging, Boosting, Sampling Methods, SVM, RF

Tensorflow, sklearn, numpy, pandas, Pytorch, TersorRT

Languages

Python, C, Java

Visualisation Softwares

Qlik Sense, Tableau, Power BI