Writabrata Dey

Data Analyst, DSP Asset Managers Pvt. Ltd. , India

writabratadey@gmail.com | 91-8967953770 | LinkedIn | GitHub

Education

Specialisation in Machine Learning & Deep Learning

2022 - 2023

Scaler Data Science, India

Modules :

- Deep Learning, Neural Network basics, CNN, Computer Vision and Natural Language Processing. Machine Learning (Supervised and Unsupervised) and Statistical Tests both from a mathematical as well as a programming point of view.
- Advanced SQL, Python, DSA basics, Git, Containerisation, CI/CD Pipeline and AWS Basics.
- Visualization tools like Tableau, Power Bi

Projects and Business Case Studies :

- **Driver Attrition in Cab Booking Industry :** Investigated and analysed the root cause of the increasing issue of driver attrition in one of India's cab booking companies. Multiple insights along with 6 to 7 key pointers were discussed post analysis. I tested multiple classification models like XGBoost, Stochastic Gradient Descent and bagging techniques like Random Forest with multiple different hyperparameters. As the data was highly imbalanced, comparing multiple evaluation metrics like Confusion Matrix, ROC-AUC and others between Class Weights and SMOTE, I went with Class weights and built the final model.
- **Loan Approval Prediction Model for a Bank :** I built an underwriting layer to determine the creditworthiness of MSMEs as well as individuals based on a detailed feature engineering and analysis of 27 features. I tested multiple classification models like logistic regressions and boosting techniques. Keeping the business case and the use case in mind, a detailed focus was given to Precision Recall while evaluating the models.

Master of Science in Physics (Spec. Astrophysics)

2018 - 2020

St. Xavier's College, Kolkata, India

Modules :

- Scientific Computation and Data Analysis and Visualisation using Python. Elements of Interfacing, Linux Command Line Interface, MATLAB.
- Superconductors, Solid State & Semiconductor Physics, Boolean Algebra, Electronic equipment like Diodes, Transistors, Microprocessors etc from both theoretical as well as experimental point of view. Advanced mathematical tools like Tensor Algebra, Higher order Differential Equations and solutions, and Group Theory.
- Detailed modules on Special and General Theory of Relativity, Particle Physics, Cosmology, Computational Astronomy.

Projects and Presentations :

- **Black Hole Singularity Replacement :** The singularity at the centre of a Black Hole due to intense space-time curvature, has been an enigma for ages. Though General Relativity fails to hold in these scenarios, there are multiple probable solutions out there. In the paper, I analysed and discussed the de Sitter solution, one of the potential solutions to address the singularity issue.
  [Paper Link]
- **Bardeen Black Hole Overview :** The Bardeen Black Hole solution is a solution without the central singularity. In this paper, I covered and explained the horizon structure and Ergosphere behaviour of a rotating Bardeen Black Hole.
  [Paper Link]
- **All India Mega Science Projects Exhibition - Vigyan Samagam :** I worked with a team of 5 students and we exhibited the Project - SKA (Square Kilometer Array) with demo models and detailed posters in this exhibition for 7 days.

Bachelor of Science in Physics

2015 - 2018

Ramakrishna Mission Vivekananda Centenery College, Kolkata, India

Modules :

- Basics of Computer Architecture & programming. Numerical Computations using C.
- Boolean Algebra and Electronics. Diodes and Transistors, Counters, Registers, Decoders, Encoders, Amplifiers.
- Linear Algebra basics with Vector Calculus, Curvilinear Coordinates, Dirac-delta. Matrix analysis, using Eigen Values and Eigen Vectors. Probability Theory including multiple Distribution Functions and Statistical tests.

Workshops :

- I attended a whole day workshop at VECC (Variable Energy Cyclotron Center), Kolkata, India to study and observe the theoretical as well as experimental part of the Proton beam behaviour and the collision consequences.

Research Interest

- **Computer Vision** applications in **Aerial Robotics**. Applications in **vision-based flight control (Autonomous Spacecraft Navigation), spacecraft smart landing** and other applications in **mission designing** and **mission autonomy**. Applications in telemetry data analysis and spacecraft health monitoring.
- **Satellite imagery** for Earth Science and Remote Sensing as well as **deep space object detection**.
- Computer Vision and Deep Learning application in **Cognitive Science** and Human Behaviour.

Personal Projects

- **Satellite Imagery Segmentation :** *[GitHub]*
- I am working on a U-NET Encoder-Decoder CNN model to train on the dataset in order to accomplish a semantic segmentation of 6 classes in this Geospatial data. These Raster satellite data from the Mohammed Bin Rashid Space Center, Dubai, contains aerial imagery of Dubai. I generated the data in the required format from the Ground Truth and Masked Images. The U-NET is being trained on the data and I am comparing the layer architecture with the help of model visualizer tools like NETRON to get a better understanding of the underlying behaviour. Post training, I will be evaluating the neural network with metrics like the Jaccard Coefficient and the best-suited Loss functions. Finally, I will try to create an automated pipeline using GitHub Action as a CI/CD pipeline tool to AWS and also to HuggingFace using Gradio.
- **Hand Gesture Volume Controller :** *[GitHub]*
- I created a pipeline which enables a user to control the volume of the computer system by hand gesture. 2 libraries are used primarily. I have used the library MediaPipe by Google in order to deal with the hand detection, more specifically the palm detection part of the project, OpenCV was used to interact with the system camera driver and also to modify the visual. As I have created the project on a personal Linux Machine (running Ubuntu 22.04.4 LTS), I used Python's Subprocess to call a bash script to interact with the audio driver. In Windows OS, the process to interact with the audio driver may vary and the Python library, pycaw can be used.
- **Face Mask Detection :** *[GitHub]*
- A pipeline to detect if the face mask is being used by the user or not. I have used the OpenCV module to interact with the system camera module and trained a VGG16 as a neural network by Transfer Learning. In the end, the model performance was evaluated and got a score of around 95% on the validation data. The dataset, the model was trained on was published by pyimagesearch.com

Experience

Data Analyst *2023 - Present*

DSP Asset Managers Pvt. Ltd., Mumbai, India

- I have built potential segments of customers with multiple Unsupervised Machine Learning techniques like KNN, DBSCAN along with the business domain knowledge. For segmentation prediction use cases, I have built and been maintaining classification models using XGBoost, AdaBoost and other boosting and ensemble techniques. The models are being retrained on a quarterly and half-yearly basis depending upon the use cases.
- I have automated multiple database updations using AWS Batch, Lambda and then AWS Eventbridge to set up the cron. I have used s3 tables and connected to Athena for further data manipulations. I also have the responsibility to assist the Data Engineering team in case of occasional data pipeline failures.
- I am looking into multiple monthly and quarterly insights, campaign evaluation decks and sudden time-sensitive data requests from multiple internal teams. AWS QuickSight and S3 are being used on a day-to-day basis in these regards.

Data Science Engineer Intern *2023*

ResoluteAI Software, Bengaluru, India

- I lead a team to build a Predictive Control Tower Dashboard. I built the Performance Monitoring System using Power BI to monitor multiple assets (system of electronic devices) installed at multiple commercial establishments across India.
- Multiple Machine Learning Models were trained to predict the health of a particular asset and then merged with Power BI to create a predictive dashboard.
- Using OpenAI API and Langchain framework in the back end and Flask and Streamlit for the front end part, I built a custom document-based conversational UI. Based on prompt vectorization and the already present vector embedding database, the pipeline was able to provide relevant response from the uploaded document. Semantic Search was playing a crucial role in the backend. Using a similar architecture, I also built a document filtration pipeline.
- Using OpenAI API, I have built a pipeline to prompt base autonomous content-creating system. Text-to-Text and Text-to-Image models were used in this regard.

Machine Learning Intern *2022*

Suvidha Foundation, Nagpur, India

- Worked on a text summarizer using Natural Language Processing. Studied and analysed research papers regarding ML.

Supply Chain Analytics Intern *2022*

Flipkart, India

- Worked on Data Collection and Analysis in the Quality Assurance department along with the Operations team to improve existing workflow and system.
- I worked with the analytics team to monitor and analyse customer orders' journey across multiple departments from the very beginning of Inbound to the Outbound and then Outbound to eventually the customers.

Skills

- Python (Data Science Libraries, OOPS), SQL, C++ , Flask, Streamlit
- AWS (QuickSight,Lambda, Batch, EventBridge, Athena), Linux Basics, CI/CD Pipeline, Docker, GitHub Action, Heroku
- SQL(Snowflake, MySQL, Postgres), Power Bi, Tableau, Git
- Machine Learning (Supervised & Unsupervised), CNN Langchain, LLM.
- Geospatial Data Analysis, QGIS, Maple, LateX