Sayan Chatterjee

csayan100692@gmail.com
Contact No. +918250232644
LinkedIn : linkedin.com/in/sayan-chatterjee-884aa138
GitHub : https://github.com/chatterjeesayan1992

August 27, 2021

Education

- **Indian Statistical Institute-M.Tech in Computer Science (2016 - 2018)**
  - Percentage of Marks-**66.67 %**
- **Jadavpur University-BE in Electrical Engineering (2010 - 2014)**
  - CGPA-**7.42**
- **Krishnagar Collegiate School-Higher Secondary (2008 - 2010))**
  - Percentage of Marks-**86.8%**
- **Krishnagar Collegiate School-Class X(2008))**
  - Percentage of Marks-**95.75%**

Industry Experiences

Data Scientist-Envestnet Yodlee (April 2020 - Present)

- **Forecasting Partially Built Time Series Data**: Mentored an intern in solving the problem of forecasting partially built time series data of several merchants in order to remove some unexpected dips in them due to lag issues.
- **Hierarchical and Grouped Time Series Forecasting**: Worked on a problem of forecasting multiple time series data with a hierarchical structure.
- **User Quality Scoring to Select Best User Cohorts for Optimization of Certain Financial Metrics**: This work required developing multiple statistical metrics to select users based on them in order to optimize certain financial metrics for a huge volume of transaction data. **Got awarded for this project along with a couple of team members**

Research Engineer-American Express ML&AI Labs (June 2019 - March 2020)

- **GAN for Minority Oversampling in Structured Financial Data**: The algorithm solves the problem of generating meaningful structured data from a very small number of structured data samples so that generated data improves the generalization performance of a downstream classification task. The algorithm had been tested in several internal domains and worked successfully.
- **Got rewarded for creating a tutorial article for a popular online machine learning tutorial Site**

- Worked on a couple of other projects regarding optimal preprocessing for training a neural network and automated extraction of useful information from the Google search result.

Associate-PwC - Data & Analytics (July 2018-May 2019)

- Built up a voice enabled SQL bot integrated with CXO dashboard for a manufacturing client.

Recent Kaggle Participations

- **Tabular PlayGround Series June - 2021**- Came in top 3% (26th among 1170 teams) in final leaderboard. The challenge was to work on a classification problem on a given tabular dataset with 75 features and 2 lac rows .

- **CommonLit Readability Prize**- This ongoing competition needs one to predict the readability score of a given paragraph based on a training data of previously scored paragraphs. Trying to create an architecture based on RoBERTA for this problem. Current score in public LB is 0.469. (among top 700 in about 2500 teams).

- **BirdClef-2021- Birdcall Identification by Cornell Lab of Ornithology** The task is to identify birds from their sounds in some long audio files containing sounds of several birds. It was a multilabel multiclass classification problem where the training data consisted of many short audio files each containing one/more than one bird sounds.

Projects and Internships

1. **M.Tech Dissertation:**
   Thesis Title : Severity Gradation of Psoriatic Plaques Using Ensemble of Deep Convolutional Neural Networks
   Area : Deep Learning in Medical Imaging
   Guide : Prof. Utpal Garain, CVPR Unit, ISI
   Remarks : The objective was to grade the severity of an immune mediated, inflammatory skin disease named Psoriasis based on three parameters namely **Erythema** (extent of redness of the diseased area), **Induration** (extent of elevation of the diseased area), and **Scaling** (extent of scaliness of the skin) using an image dataset collected in an uncontrolled environment by layman photographers. The approach gave an accuracy improvement of **7-8%** from that of an approach tried in a paper published in **ICPR,2016**

2. **Research Internship (IIT Kharagpur):**
   Project Title : Anomaly Detection of Time Series Data using Fuzzy C-Means Clustering and LSTM
   Area : Deep Learning in Time Series Analysis
   Guide : Prof. Sudeshna Sarkar, HOD, CSE Department, IIT Kharagpur

3. **Project with TCS-iON and ISI:**
   Project Title : Form and Performance Analytics of a Large Scale Online Examination

Area : Statistical Data Analysis
Guide : Prof Nachiketa Chattopadhyay (Head, SOSU Unit, ISI Kolkata), Dr. Asit Chakraborty (Prof, SOSU Unit, ISI Kolkata), Dr. Pinakpani Pal (Prof. ECSU Unit, ISI Kolkata)

4. **Cognitive Science Project:**
Project Title : Effect of Constructive Video Games on Working Memory
Area : Cognitive Science
Guide : Prof Garga Chattopadhyay (Prof, PRU Unit, ISI Kolkata).

Skills

- **Programming and Scripting Languages:** Python, C, C++, Unix Shell, R, GNU Octave
- **Parallel Programming and Cloud Platform**: PySpark, QSUB (SGE Cluster), AWS (Redshift, EC2, EMR, S3)
- **Deep Learning Frameworks:** Tensorflow, Pytorch, Keras, Trax

Relevant Courses Gone Through:

Deep Learning for Computer Vision, Pattern Recognition and Image Processing, Natural Language Processing, Cognitive Science, Probability and Stochastic Processes, Elements of Algebraic Structures, Advanced Digital Signal Processing

Awards & Achievements

- Achieved a rank of **236 (among 127196 students)** in **West Bengal Joint Entrance Examination, 2010**.
- Owner of a scholarship for outstanding performance in 10+2 level board examination, 2010 from Ministry of Human Resource Development, Govt. of India.
- Achieved a state rank of **729 (among a million students)** in **All India Engineering Entrance Examination**, 2010.

Certifications

- **Natural Language Processing by Attention Models**- DeepLearning.AI (Coursera)
- **Advanced Computer Vision with Tensorflow**- DeepLearning.AI (Coursera)

Hobbies

- Soft Pastel Painting, Portrait Sketching/Painting (won a couple of prizes), Reading Books