

SARTHAK CHAKRABORTY

+91-9836560275 | sarchakr@adobe.com | sarthak-chakraborty | sarthak-chakraborty.github.io

INTERESTS

Data-driven Systems, ML for Systems, Edge Computing, Distributed Systems, Machine Learning

EDUCATION

- **Dual Degree (B. Tech + M. Tech) in Computer Science and Engineering** July 2016 - April 2021
Indian Institute of Technology Kharagpur, India
Cumulative GPA: 9.74/10.00 (Class Rank 2)

PUBLICATIONS

- [1] Lovish Chopra*, Sarthak Chakraborty*, Abhijit Mondal, and Sandip Chakraborty. **PARIMA: Viewport Adaptive 360-Degree Video Streaming**. In Proceedings of the Web Conference 2021 (WWW '21), April 19-23, 2021, Ljubljana, Slovenia [LINK]

WORK EXPERIENCE

- **Research Associate - Adobe Inc. (Big Data Experience Lab)** Jul 2021 - ongoing
Group: Data-driven Systems, Insights and Experience *Bangalore, India*
 - * Joined the Big Data Experience Lab at Adobe India under the Data & Systems group and will be collaborating with multiple researchers from the group
- **Research Intern - Adobe Inc. (Big Data Experience Lab)** Apr 2020 - Jul 2020
Topic: Architecting Asynchronous Federated Learning *Sunav Choudhary, Manoj Ghuhan*
 - * Designed a scalable, flexible, robust and distributed framework for federated learning that supports synchronous as well as asynchronous modes of model training
 - * Supported On-Client training in federated fashion on various target devices including android mobiles (tfLite), web browsers (tfjs), IoT (Raspberry Pie) and desktop
 - * Devised an algorithmic strategy to effectively aggregate stale gradients and deployed the framework on over 100 clients to perform image classification and image segmentation task with real world production model
- **MITACS Globalink Research Intern - University of Waterloo** May 2019 - Aug 2019
Topic: Advanced Optimization Methods for Machine Learning *Dr. Hans de Sterck*
 - * Designed a randomized ALS algorithm targeted for CP Decomposition and Completion of Sparse Tensors
 - * Computed leverage scores of the rows of factor matrices to sample observed non-zero data points using weighted reservoir sampling and Partial Sum Tree based sampling strategies
 - * Performed several diagnostics and validated our method against benchmark algorithms like conventional ALS, SGD, CCD++ and RRALS algorithms
<https://github.com/sarthak-chakraborty/rrals>
- **Undergraduate Research Intern - IIT Kharagpur (Funded by Shell India Pvt. Ltd.)** May 2018 - Oct 2018
Topic: Unsupervised Clustering and Estimation of Model Parameters using GMM *Dr. Swanand Khare*
 - * Designed and implemented a randomized EM algorithm to solve the unsupervised clustering problem
 - * Modelled the data in Gaussian Mixture Model framework to estimate its parameters by introducing randomization in between successive EM steps
 - * Tested the effectiveness of the algorithm against standard approaches like Lloyd's algorithm, SEM, CEM using a set of diverse synthetic, real and industrial datasets
<https://github.com/sarthak-chakraborty/Estimation-of-Model-Parameters-using-GMM>

PROJECTS

- **Cross-Chain Training of Learning Models via Blockchain Interoperability** Aug 2020 - Apr 2021
Dr. Sandip Chakraborty *Master's Thesis Project*

 - * Developed an end-to-end system for training a model and transfer of model state over two permissioned blockchain networks via the concepts of interoperability to facilitate transfer learning
 - * Designed a synchronous Federated Learning system to train models on multiple physical devices as clients
 - * Incorporated a permissioned blockchain network to store the state of the model learned by the federated system such that it can provide enough information to make the state auditable
 - * Constructed a relay-based cross-chain transfer mechanism to transfer the model state from one network to the other via HTTP channel. Signatures ensured the verifiability and authenticity of the data transferred.

- **PARIMA: Viewport Adaptive 360-degree Video Streaming** Jul 2019 - May 2020
Dr. Sandip Chakraborty *Bachelor's Thesis Project*

 - * Designed an online viewport adaptive video streaming algorithm along with a client-server streaming platform
 - * Developed a novel PARIMA algorithm: an augmented Passive-Aggressive(PA) model and time series(ARIMA) model for viewport detection using video content as well as personalized head movement tracking.
 - * Employed a pyramidal adaptive bitrate allocation scheme to maximize the Quality of Experience
 - * Used HEVC video encoding, GPAC for segmenting video chunks and 'MP4Client' for client streaming of video
<https://github.com/sarthak-chakraborty/PARIMA>

- **Scalable Method for Representing Large Scale Graphs** Aug 2019 - Jan 2020
Dr. Sourangshu Bhattacharya

 - * Developed a hierarchical community-detection based algorithm for network embedding of large scale graphs
 - * Constructed hierarchy tree using Louvain community detection algorithm and studied the community structure of the graph to establish relevant inter-community links at each hierarchy level
 - * Generated embedding using Node2vec/Deepwalk at each hierarchy level and combined the individual node embeddings to get the network embedding

- **Minimally Supervised Semi-Supervised Text Classification** Jan 2019 - Oct 2019
Dr. Jiaul Hoque Paik

 - * Designed an algorithm aimed to select a minimal set of samples for semi-supervised learning and achieve par accuracy compared to a fully supervised model
 - * Obtained semantic vectors using self-attention based bidirectional LSTM network and generated topology adaptive hyper-cuboids using bisection based homogeneity reduction algorithm
 - * Constructed similarity kernel and used k-DPP to select diverse set of samples from each hyper-cuboid
<https://github.com/sarthak-chakraborty/MinSSL>

- **Question Answering over Linked Data (QALD)** Aug 2018 - Nov 2018
Dr. Plaban Kumar Bhowmick

 - * Devised an algorithm to translate natural language query into SPARQL query and retrieve answer
 - * Analysed the natural language query to determine entities and candidate predicates by eliminating stopwords
 - * Explored various NLP based tools like SpaCy and CoreNLP to generate dependency graph, which was then used along with POS tagging in a heuristic template based algorithm to obtain the SPARQL query
<https://github.com/sarthak-chakraborty/QALD>

- **Personal Library System** Jan 2018 - Apr 2018
Dr. Sudip Misra

 - * Developed a GUI based software using JAVA Swing and MySQL to automate the proceedings of a library
 - * Incorporated features which helps the owner issue and update book information, check availability of each book, and let users borrow books

- * Implemented Waterfall SDLC model along with industry level software development techniques like SRS, DFD and UML for designing and JUnit for testing
<https://github.com/sarthak-chakraborty/Personal-Library-System>

TEACHING EXPERIENCE

- Undergraduate Teaching Assistant for *Database Management Systems (CS43002)* course at IIT Kharagpur 2021
- Undergraduate Teaching Assistant for *Theory of Computation (CS41001)* course at IIT Kharagpur 2020

CERTIFIED WORKSHOPS

- **Google India AI Summer School** Aug 2020
Google Research India
 - * Successfully completed the three-day long summer school on AI organized by Google Research India
 - * Attended graduate level Machine Learning lectures from experts around the world and participated in discussion forums with leading Google researchers from India and around the world
- **Image Processing Workshop** Dec 2016
IEEE Robotics Winter Workshop
 - * Successfully completed the week-long IEEE certified winter workshop on Image Processing at IIT Kharagpur.
 - * Applied various image processing techniques using OpenCV library to implement movement detection algorithm based on background subtraction and colour extraction.

MISCELLANEOUS PROJECTS

- Developed a Distributed Collaboration System where multiple users can collaborate on a single document at once that maintained consistency along with a passive replication scheme. It used a master-worker architecture of servers.
- Implemented Multimodal Emotion Classifier for conversations in MELD datasets. Used visual, textual, and acoustic as individual modes and trained a trimodal classifier to combine the activations from these modes
- Implemented REINFORCE policy-gradient algorithm to maximize rewards in the Banana environment of Unity
- Implemented a SPAM/HAM classifier using neural networks and Decision Tree based classifier from scratch
- Developed a web-based and an android-based portal linked with SQL for the purpose of Bicycle Renting
- MRP: Implemented a reliable message-oriented communication protocol over an unreliable User Datagram protocol
- Developed APIs of memory-resident file systems for linked-list based FAT and indexed based inode implementations
- TinyC: Implemented a compiler for a subset of C functionalities to translate the C code to x86 Assembly Language
- KGP-RISC: Designed a 32-bit single cycle CPU(RISC based architecture) in Verilog VHDL and tested it on FPGA

SKILLS

- | | |
|----------------------------------|--|
| • Languages | Python, C, C++, Java, SQL, Golang, Verilog, MIPS, Scala |
| • Packages and Frameworks | scikit-learn, C++ STL, Git, Keras, tensorflow, creme, PyTorch, OpenCV, Hyperledger Fabric, MongoDB, Docker, Kafka, NLTK, Spark |
| • Web Development | HTML, CSS, PHP |

SCHOLASTIC ACHIEVEMENTS

- Graduated with a Department Rank and an Institute Rank of 2 among all the Dual degree students 2020
- Recipient of the Goralal Syngal Memorial Scholarship awarded by the Institute for academic excellence. 2020, '19
- Acknowledged by the Department of Computer Science and Engineering for performance par excellence. 2018

- Secured a Department change from Ocean Engineering to Computer Science and Engineering due to excellent academic records in the fresher year 2017
- Within top 1% among 2 lakh students in JEE(Advanced) and 0.5% among 1 lakh students in WBJEE 2016

RELEVANT COURSES

- **Core Courses:**

Programming and Data Structures[#], Algorithms-I[#]&II, Discrete Structures, Formal Language and Automata Theory, Switching Circuits and Logic Design[#], Software Engineering[#], Compilers[#], Computer Organization and Architecture[#], Operating Systems[#], Computer Networks[#], Database Management Systems[#], Theory of Computation, High Performance Computer Architecture, Machine Learning, Artificial Intelligence, Scalable Data Mining, Reinforcement Learning, Deep Learning, Advances in Operating Systems Design, Advanced Database Systems⁰ Distributed Systems, Social Computing, Principles of Programming Languages [# Practicals Involved]

- **Other Courses:**

Maths I & II, Probability and Statistics, Linear Algebra, Signals and Network, Knowledge Modelling and Semantic Technologies, Educational Data Analytics, Regression and Time Series Models, Operations Research

EXTRA-CURRICULAR

- Presented a talk on 'Predictive Maintenance' at the Energy Transition Technology Summit (ETTS), 2018, held at Shell Technology Center, Bangalore 2018
- Had been a member of Team AUV, IIT Kharagpur and worked on design changes to bring stability in the existing model of underwater vehicle, Kraken 3.0 2017-2018
- Had been an active member of the Aquatics Society as a part of National Sports Organisation 2016-2018
- Secured gold in Open IIT Sketching competition held in my fresher year 2016
- Completed Bhushan (3 year diploma course) and Visharad (2 year diploma course) in Fine Arts under Pracheen Kala Kendra, Chandigarh 2008-2013