Prateek Chanda

Computer Science PhD Student, IIT Bombay - Prime Minister Research Fellow

prateekkol21@gmail.com || prateekchanda@google.com

prateekiiest.github.io | Google Scholar | ResearchGate | GitHub | LinkedIn | +91-8337055526

EXPERIENCE

Google Research India	**Machine Learning Optimization & Cognitive Machine Learning**	Bangalore , India
Student Researcher	Manager : Pradeep Shenoy	July 2023 - Present

- Building model dependent and data distillation techniques for pretraining efficiency in Large Language Models, ongoing experiments using Google LaMDA models
- Implemented Matformer Paper on top of Google LaMDA models and incorporated data distillation techniques and gained better performance of downstream tasks +0.4% gains on GPT Gen and +0.57% gains on GPT Rank Tasks as compared to vanilla training
- Currently looking into implementation of effective learning strategies like curriculum learning/ reweighting training instances via Meta learning for better LLM pretraining efficiency and how effective skill ordered training Skill-It paper can help achieve better.

Microsoft Research India	**Center for Societal impact through Cloud and Artificial Intelligence**	Bangalore , India
SCAI Software Engineer	Supervisor : Dr. Amit Sharma	Nov 2019 - Feb 2022

- Developed user feature embeddings based on user responses logging and telemetry logging for building recommendation models for recommending different therapeutic activities and sections of a mental health application as interventions to the user.
- Started with initial collaborative and content based filtering for recommendations, and further developed a causal recommendation model where each micro-intervention was indicated as treatment to the user.
- Shipped full end-to-end product that captures patient mental health expert conversations along with different modules for suitable patient interventions. (Papers accepted at *ACM CSCW*)

Indian Statistical Institute	**Indian Statistical Institute**	Baranagar , India
Research Assistant (Remote)	Advisor : Malay Bhattacharyya	May 2021 - Present

- Performed EHR analysis based on Sage Bio Networks Competition to detect which patients are most Covid vulnerable.
- Employed models like Graph Neural Networks, InterpretML, and other ensemble classification algorithms. Report | Models Explored
- Proposed a distributed anomaly detection frame work in large data streams in an online manner. *Currently under review at VLDB*

Machine Intelligence Unit	**Indian Statistical Institute India**	Baranagar , India
Research Intern May '18	Advisor : Dr. Ashish Ghosh	Project Report	Code

- Performed theoretical study on different Metric Learning algorithms to learn similarity metric from data distribution.
- Did an empirical analysis as well as evaluation of metric learning methodologies w.r.t different datasets like Iris, Wine Dataset, thus showcasing performance & limitations across various data distribution.

SunPy	**NASA Open Source Software	Solar Data Analysis in Python**	Remote, US
Google Summer of Code	Technologies : Python , Git	Dec 2016 - Apr 2018

- Collaborated with a team of 60 researchers to develop modules for efficient solar data retrieval, data processing and storage functionality for data analysis.
- Implemented a solar data retrieval system to collect solar data from solar observatories based on date in an SQL data base toanalyze different helio-features from the data over a period of 10 years. Used by the SunPy project.
- Implemented proposed solar image processing algorithm from research paper achieving 18% improved memory utilization and better feature extraction with less noise. Got acknowledged along with researchers at NASA Goddard Space Flight Center in nine Software Releases.

PROJECTS   MORE PROJECTS ON GITHUB : PRATEEKIIEST/REPOSITORIES

IITB-Bayesian Coresets for Personalized Federated Learning   *Course CS 769 Project: Prof. Ganesh Ramakrishnan*
- As part of CS 769 : Optimization in Machine Learning, I worked with Prof. Ganesh on combining coreset based subset selection strategies for a personalized federated learning setting where each client only trains on a subset of their individual data, resulting in about near-optimal performance had they trained on their individual original data. **Under Review at ICLR 2024**

IITB-SketchSimRank - SimRank for Graph Streams (Theoretical Guarantees) *Course CS 635 Project: Prof. Soumen Chakrabarti*
- As part of the CS 635 Project, we propose simrank with approximate guarantees for graph streams (dynamic graphs) using only fixed space via hashing (count min sketch). We achieve theoretical guarantees on approximate simrank computation.

Future Promising Heavy Hitter Detection in Streaming environments Advisor : Bryan Hooi, NUS (2020 - July 2021)
- Proposed a novel anomaly scoring technique for the application of heavy hitters using apache datasketches frequent sketches and cumulative distribution comparisons based on quantile sketches. **-** Performed statistical tests to validate the accuracy of sketches for large dataset items in our anomaly scoring function via sketch guarantees and confidence bound.

PUBLICATIONS

- **Bayesian Coreset Optimization for Personalized Federated Learning** : Under Review ICLR 2024   (Avg Rating: 6.33)
  Prateek Chanda & *Shrey Modi* & *Ganesh Ramakrishnan*

- **Would I regret later joining this Community ? Using temporal neighborhood information for community retention in a game theoretic community detection framework** : AAAI-22 Workshop on Machine Learning for Operations Research (ML4OR)
  Prateek Chanda & *Susanta Chakraborty*

- **Robust Deep Reinforcement Learning Control framework** : AAAI-22 Workshop on Robust Artificial Intelligence System Assurance
  Prateek Chanda

- **MINDNOTES: A Mobile Platform to enable users to break stigma around mental health and connect with therapists**: CSCW Demo 2021
  Prateek Chanda, Amogh Wagh, Jemimah A Johnson, Swaraj Renghe, Vageesh Chandramouli, George Mathews, Sapna Behar, Poornima Bhola, Girish Rao, Paulomi Sudhir, TK Srikanth, Amit Sharma, Seema Mehrotra

- **Reaching out : Towards a sustainable allocation strategy between users and therapists**: Accepted NeurIPS Machine Learning in Public Health
  Prateek Chanda

- **Distributed Anomaly Detection in Edge Streams using Frequency based Sketch Datastructures**: arxiv
  Prateek Chanda, Malay Bhattacharyya

- **A Sketch Based Game Theoretic Approach to Detect Anomalous Dense Sub-Communities in Large Data Streams**: arxiv
  Prateek Chanda, Aadirupa Saha

- **A Novel Graph Based Clustering Approach to Document Topic Modeling**: Accepted *9th ICCCNT 2018, IISc*
  Prateek Chanda, Asit Kr Das
- **SunPy A Python package for Solar Physics**: Journal Paper *Journal of Open Source Software 2020*
  *Stuart Mumford, **Prateek Chanda**, The SunPy Community*
- **SunPy v1. 0, the community-developed, free and open-source solar data analysis environment for Python.** : Journal Paper *American Geophysical Unit / NASA ADS 2019*
  *Stuart Mumford, **Prateek Chanda**, The SunPy Community*

- **The sunpy project: Open source development and status of the version 1.0 core package**: *The Astrophysical Journal 2020, IOP*
  *Stuart Mumford, **Prateek Chanda**, The SunPy Community*

ACHIEVEMENTS

- **AISTATS 2022 Mentorship Program**: Selected for AISTATS Mentorship Program to work with Aadirupa Saha at Microsoft Research NYC
- **AI for Science NeurIPS Workshop Mentorship Program**: Selected for AI for Science NeurIPS Mentorship Program to work with Malay Bhattacharyya at Indian Statistical Institute
- **Microsoft Garage Hackathon 2020**: Recipient of Hackathon 2020 NGO award from Microsoft Garage India under AI for Social Good.
- **Microsoft Research India Sponsorship**: Recipient of MSR India Sponsorship Funding for internship work at IIT Kharagpur
- **GAABESU research award IIEST**: Received GAABESU(IIEST) research award for research contributions for academic year 2018
- **JBNSTS Scholar**: Selected for Jagadis Bose National Science Talent Search Scholarship
- **AIEEE Merit**: Within top 0.26% of applicants in All India Engineering Entrance Exam approx. 1.3 million people

SKILLS

- **Languages**: Python, C++, SQL, Java, C#, TypeScript   **Technologies**: Azure, Azure ML Studio, GitHub, GitLab, Jekyll, GCP
- **Libraries**: TensorFlow, PyTorch, Scikit-Learn, Pandas, Jupyter, Microsoft Graph SDK

EDUCATION

Indian Institute of Technology, Bombay   Maharashtra, India
 Ph.D. in Computer Science & Engineering (Generously funded by **Prime Minister Research Fellowship**)   2022 - Present

   Advised by **Prof. Ganesh Ramakrishnan**

Relevant Graduate Coursework: CS 769: Optimization in Machine Learning (**AA**), CS 726: Advanced Machine Learning (**AB**), CS 725: Foundations of Machine Learning (**AB**), CS 635: Information Retrieval & Mining for Hypertext & the Web (**BB**), IE 643 Deep Learning and Theory (**AB**), CS 802 Seminar (**AA**)
Teaching Assistantship: CS 335 + CS 337 : Artificial Intelligence & Machine Learning Course and Lab (Fall 2022)
   CS 419M : Introduction to Machine Learning (Spring 2023)

Indian Institute of Engineering Science and Technology, Shibpur   Howrah, India
 Bachelor of Technology in Computer Science & Engineering;   *First Class Honors*   GPA: 8.86/10.0   WES: 10/10   2015 - 2019

Thesis: *Avoiding Past Choice Regrets: A Game Theoretic Community Detection using Temporal Information*
   Advisor: Malay Kule & Dr.Susanta Chakraborty   Thesis Report

Relevant Coursework: Data Structures & Algorithms, Operating Systems, Database Management, Cloud Computing & Big Data, Machine Learning & AI, Probability & Statistics, Discrete Structures, Computer Graphics,Computer Networks, Computer Architecture

POSITIONS OF RESPONSIBILITY

- **Program Committee for ODD SIGKDD workshop** 2021 - Workshop Link
- **Reviewer** for IEEE Transactions on Mobile Computing , COMSNETS, AISTATS, Journal of Open Source Software
- **Google Code In**, **GSoC Mentor** : Mentored over 80 students under Google Code In 2018, Hacktoberfest 2018, 2017
- Leading the open source club at Campus as a GitHub Campus Expert - organising hackathons and open source mentorship programs in campus and engaged students from different departments in open source