# Mayank Kakodkar

Purdue University, West Lafayette, IN 47906 USA

mayank.kakodkar@gmail.com | https://www.linkedin.com/in/mayankkakodkar | https://kakodkar.github.io/

## Education

**Purdue University, West Lafayette, IN, USA**                                        August 2016 - May 2022 *Expected*
*Doctor of Philosophy in Computer Science*, GPA: 4/4
Advisor : Prof. Bruno Ribeiro, Research topic : Regeneration in discrete state and reversible Markov chain Monte Carlo

**Indian Institute of Technology - Delhi, India**                                        August 2007 - June 2011
*Bachelor of Technology, Mechanical Engineering with a minor in Computer Science and Engineering*

## Publications

**M. Kakodkar** and B. Ribeiro, "Sampling Subgraphs via Regenerative Tree Sampling: " (Working Title)

C. Teixeira[*], **M. Kakodkar**[*], V. Dias, W. Meira and B. Ribeiro, "Sequential Stratified Regeneration: MCMC for Large State Spaces with an Application to Subgraph Count Estimation" (Accepted) Data Mining and Knowledge Discovery, 2021, Preprint on Arxiv

T. Ober, **M. Kakodkar**, Y. Cheng, P. Gonsalves, P. Brenner, E. Johnson, B. Ribeiro, J. Zdankus, "Curating Computer Science Educational Content with Machine Learning" (Accepted) APA Technology, Mind & Society, 2021.

N. Linnel, A. Dayal, P. Gonsalves, **M. Kakodkar**, B. Ribiero, A. Starr, T. Urdan, J. Zdankus, "Curated Pathways to Innovation: Personalized CS Education to Promote Diversity" In the Journal of Computing Sciences in Colleges, 2020

P. Savarese, **M. Kakodkar**, and B. Ribeiro, "From Monte Carlo to Las Vegas: Improving Restricted Boltzmann Machine Training Through Stopping Sets" In the Thirty-Second AAAI Conference on Artificial Intelligence, 2018 (AAAI-18)

## Work Experience

**Research Assistant (ML Dev/Researcher)**                                        January 2017 - Present
*Curated Pathways to Innovation(CPI), Headquartered at San Jose, CA, USA*
*Purdue University, West Lafayette, IN, USA*
Solely responsible for improving and maintaining the recommendation algorithm, continuous AB testing and reporting applications. Co-ordinated with teams in HPE, Notre Dame, SCU and YWCA-SV to provide a web-based "virtual guidance counselor" to schools and organizations. CPI uses gamification to guide and incentivize students to take up Computer Science and STEM as a profession.

**Applied Science Intern (L5)**                                        May 2021 - September 2021
*Amazon Advertising, New York, NY, USA*
Created a semi-supervised model to detect New-to-Brand and Attrition events in user-brand relationships. Proposed simplified, (spark) scalable models which passed multiple pre-defined goodness score tests. The project was appreciated for being almost production-ready within a short time frame and for garnering interest from sister teams to make the product live.

**Graduate Teaching Assistant**                                        August 2016 - December 2016
*Purdue University, West Lafayette, IN, USA*
Undergraduate Data-structures and Algorithms with Prof. Susanne Hambrusch
Created assignments, conducted tutorials, maintained online forums and graded.

**Module Lead - Research and Development**                                        June 2011 - June 2016
*Media.Net (Directi Group), India*
Led a team of five Java developers and two system admins in the keyword intelligence team (AutoOpt) which worked on ML and DM solutions for the contextual advertising engine.
Media.net is the second largest contextual advertising company by revenue globally with ad supply to half a million websites.

## Technical Skills

**Programming Languages:** Python (incl. Pytorch and other ML/DM/DL tech), Julia, C++, Java, R
**Other Tech:** Bash/Zsh, Slurm/PBS, (MS/PG/My)-SQL, MATLAB, Redis, Spark, Redshift, Hadoop and Hive.

## Projects

**Purdue University, West Lafayette, IN, USA**
- **Negative Sampling for Link Prediction:** Attempted to improve negative sampling through an MCMC procedure that samples negative samples proportional to the norm of the gradient.
- **Finite Mixture Model with Spatial Prior for Neighborhood-aware Regression (Course Project):** Showed that using location aware clustering based priors improves regression performance compared to a Finite Mixture Model in the task of predicting property rates.
- **A Deep Learning Approach to TribeFlow (Course Project):** Proposed using Latent Markov Embeddings to improve Tribeflow performance where user and item features are available.
- **Loan Default Prediction (Course Project):** Evaluated Support Vector Machines and Logistic Regression over the namesake Kaggle dataset.

**Media.Net - ML Specific Projects**
- **Behavioral Targeting:** Categorized visitors based on keyword click-through using a hybrid between collaborative filtering and the pagerank algorithm.
- **Sensitivity Filtering:** Replaced a manually curated system to classify and ban ads on web pages related to deaths, sexual violence, tragedies or atrocities toward children with a Naive Bayes classifier.
- **Keyword CPC Predictor:** Improved keyword CPC calculation using random forests trained on features like the time of the day, geography, DMA and exposure.
- **Attendance Intelligence (Hackathon Project):** Used Decision Trees to find effects of performance appraisals, weather conditions and national holidays on the attendance, time spent and overall productivity of employees based on scraped data from the office internal attendance portal and built a live dashboard.
- **Interest Breakout Predictor:** Predicted interest breakouts in stocks using ARIMA models trained on stock web page hits. Successfully predicted breakouts related to impending layoffs, quarterly results, restructuring, product launches and takeovers before the actual news was released.
- **System Anomaly Detector:** Built an ensemble of Decomposition based, ARIMA based and heuristic models of response times, revenue earned and API calls to auto detect root causes for system outages.
- **Fast Contextualization:** Improved the webpage contextualization engine through NER, location awareness and Bandit based optimizations. Improved the speed and scale of the ultra fast heuristic based contextualizer which computes top concepts in a page based on commercial value and relevance.
- **Keyword Relationship Generation:** Developed a bag of words based heuristic which mines keyword-keyword relationships based on co-occurrence on webpages which led to a 50% increase in revenue during AB testing.
- **DOM Classification:** Developed a fast clustering algorithm for pages of a website based on DBScan using similarities between Tries of the DOM structures as a similarity metric. Improved speed and accuracy compared to the existing system.

## Graduate Courses

Deep Learning, Bayesian Data Analysis, Data Mining, Statistical Machine Learning, Randomized Algorithms (Audit), Mathematical Toolkit For CS, Algorithm Design Analysis & Implementation, Data Communication & Computer Networks

## Service, Awards & Honors

- Reviewer ICML 2020, ICML 2021, NeurIPS 2021, ICLR 2022, SDM 2022, AAAI 2022
- Panelist for Purdue Undergraduate Research Experience, 2018, 2019
- **Ranked 559** in the **Joint Entrance Examination, IIT 2007** out of a total of more than 300,000 candidates in India.
- Awarded the **NTSE** and **CBSE** Merit Scholarships, respectively in 2005 and 2007, with acceptance rates of less than 0.5%.

## Research Presentations

**Poster presentation (Best poster award)** May 2021
*SIAM SDM Doctoral Student Forum, Remote*
**Topic:** Sequential Stratified Regeneration: MCMC for Large State Spaces with an Application to Subgraph Count Estimation

**Poster presentation** March 2020
*Prospective PhD Student Visit Day, Purdue University, West Lafayette, IN, USA*
**Topic:** From Monte Carlo to Las Vegas: Improving Restricted Boltzmann Machine Training Through Stopping Sets

**Research talk** July 2018
*[Thanes](), Inria, Sofia Antipolis, France*
**Topic:** Regeneration in discrete state and reversible Markov chain Monte Carlo

**Poster presentation** April 2018
*CERIAS Information Security Symposium, Purdue University, West Lafayette, IN, USA*
**Topic:** From Monte Carlo to Las Vegas: Improving Restricted Boltzmann Machine Training Through Stopping Sets

**Poster presentation** February 2018
*Thirty-Second AAAI Conference on Artificial Intelligence, 2018 (AAAI-18)*
**Topic:** From Monte Carlo to Las Vegas: Improving Restricted Boltzmann Machine Training Through Stopping Sets