Aniket Anand Deshmukh

CONTACT INFORMATION

Google Scholar: shorturl.at/cmxJS
Email: aniketde@umich.edu
Web: https://aniketde.github.io/
Phone: (734) 277-3529

EDUCATION

University of Michigan (UMich), Ann Arbor, MI USA — CGPA: 4.0/4.0
Ph.D., Electrical and Computer Engineering — *Aug'13 - Jan'19*

- PhD Thesis Projects: Domain Generalization [6], Multitask Learning for Contextual Bandits [11], Simple Regret Minimization for Contextual Bandits [10]
- Graduate Student Instructor: EECS 545 Machine Learning (Fall 2016, Fall 2018)

Indian Institute of Technology Hyderabad (IIT-H), India — CGPA: 8.63/10.0
Bachelor of Technology with Honors, Electrical Engineering — *Jul'09 - Jul'13*

PROFESSIONAL EXPERIENCE

Applied Scientist, AWS AI Labs, Amazon, Santa Clara, CA, USA — *Jan'23 - Present*

- LLM Pretraining for code generation:
  - Working on model based data quality for pretraining with focus on 10-20 B parameter model training
- Tech Lead, Personalizing Retrieval-Augmented Generation (RAG):
  - Led the development of a personalized RAG model for an AWS service (Amazon Q) with 10+ team members, leveraging Large Language Models (LLMs) to enhance efficiency and user experience.
  - The proposed approach achieved precise identification of personalization needs in approximately 10% of all queries, successfully boosting performance from 45% to 65% in terms of acceptability metric within these targeted scenarios in Amazon's internal use cases.
- Tech Lead, Next Best Action (NBA):
  - Led the Next Best Action (NBA) project, advancing personalization by recommending actions to boost long-term user engagement and loyalty. Developed a propensity model predicting user actions, with NBA adopted by 10+ AWS customers shortly after launch.
  - Conducted interviews with 20+ external AWS customers and 5+ Amazon teams to understand NBA needs, shaping the project's direction. Collaborated closely with product managers and a team of 10+ engineers to design, model, and launch NBA as an AWS service at Re-Invent 2023.
- Chat for Data Prep using Large Language Models (AWS Low code No Code):
  - Contributed to the "Chat for Data Prep" feature in AWS Sagemaker Canvas, utilizing LLMs for intuitive chat-based data queries, manipulation, and visualization..
  - Designed the automation for query recommendations using LLMs and created an automated testing engine to assess the "Chat for Data Prep" pipeline. Chat for Data Prep tool was launched as an AWS service at Re-Invent 2023.
- Mentor and Hiring Lead, Amazon Personalize 2024 Internships and Mentor for 2023 Interns:
  - Spearheaded the intern selection process, from outreach to finalizing hires. Successfully shortlisted over 15 candidates from hundreds of applications, conducted interviews, and coordinated the allocation of selected interns to various project teams.
  - Submitted two papers related to LLM and human feedback at NeurIPS with 2023 interns [1, 2]

Senior Applied Scientist, Microsoft Ads, Mountain View, CA, USA — *Mar'19 - present*

- Multi Media Ads (Retrieve images from general purpose corpus for Ad text):
  - Enhanced Multi Media Ads by developing an image-text multimodal model, achieving a 1.5% increase in click-through rate via A/B testing and outperforming OPEN AI's CLIP model by 9.4% in internal evaluations.
  - Coordinated across functions, leading meetings, documentation, and GPU resource requests, and collaborated with managers and engineers to establish an evaluation pipeline and A/B testing framework.

- Image attribute classification for product shopping:
  - Developed and deployed a curriculum learning-based image classification model for attribute classification on Microsoft Ads shopping platform, generating $2M daily revenue; work published in ECCV 2020 [8].
  - Labelling images for the project was expensive and time consuming. Mentored an intern on a state-of-the-art clustering method to streamline image labeling, enhancing team efficiency; findings published in IJCNN 2021 [5].
- Smart Campaigns (AI powered feature to help small businesses manage their advertising):
  - Developed a representation learning model to assess keyword-query similarity based on click data, now aiding over 10,000 advertisers on the Microsoft Ads platform in identifying relevant keywords for their ads.

Graduate Student Research Assistant, UMich, Ann Arbor, MI, USA *Aug'13 - Jan'19*

- Contextual Bandits:
  - Modeled the problem of providing personalized recommendations in the multi-task learning for contextual bandits setting and proved the regret bound [11].
  - Modelled a resource constrained, adaptive sensor selection in interplanetary spacecraft using a novel simple regret minimization framework in contextual bandits. Demonstrated improvement in simple regret of more than 25% over algorithms designed to minimize the cumulative regret [10]. This project got the multiple awards especially the best paper award at the ICML 2019 workshop of "Exploration in RL".
- Domain Generalization (Out of Distribution Learning):
  - Reduced the time complexity of kernel-based domain generalization algorithm from $O(n^2)$ to $O(n)$ using kernel approximation technique and proved the upper bound on the approximation error [6].

Applied Scientist Intern, Microsoft AI & Research, Sunnyvale, CA, USA *May'18 - Aug'18*

- Proposed and implemented Word2Vec variant of CDSSM algorithm to improve text embedding of a search query from click feedback.
- Developed self supervised learning for contextual bandits algorithms. Achieved the cumulative reward gains from 0.5% to 46.11 % over Neural-UCB on standard computer vision datasets [12].

Research Intern, Mitsubishi Electric Research Labs, Cambridge, MA, USA *May'16 - Aug'16*

- Proposed a distribution free, graph based approach for semisupervised transfer learning.
- Demonstrated that the additional information added in the form of unlabeled data improves prediction by 5.9% on the Parkinson's Telemonitoring dataset as compared to transfer learning method.

Skills

Languages	Python, SQL, Latex, HTML
Tools	PyTorch, Keras, TensorFlow, Scikit-learn, Pandas, Matlab, Azure ML, AWS, etc

Academic Service

- Outstanding Reviwer award at NeurIPS 2021, AISTATS 2022 and NeurIPS 2023 *Jun'21 - current*
- Reviewer, NeurIPS, ICML, ICLR, CVPR, AISTATS, JMLR, UAI, AAAI, etc *Jan'18 - current*
- Co-chair, TheWebConf 2023 Decision Making for IR and Recommender Systems *Oct'22 - Apr'23*
- Co-chair, ICLR 2023 Domain Generalization Workshop *Oct'22 - Apr'23*

Selected Publications

1. Subhojyoti Mukherjee, Ge Liu, Aniket Deshmukh, Anusha Lalitha, Yifei Ma, and Branislav Kveton. Experimental design for active transductive inference in large language models. *ICML submission*, 2024
2. Subhojyoti Mukherjee, Anusha Lalitha, Kousha Kalantari, Aniket Deshmukh, Ge Liu, Yifei Ma, and Branislav Kveton. Optimal design for k-way human feedback. *ICML submission*, 2024
3. Aniket Anand Deshmukh, Jayanth Reddy Regatti, Eren Manavoglu, and Urun Dogan. Representation learning for clustering via building consensus. *arXiv preprint arXiv:2105.01289. Accepted at Springer Machine Learning Journal*, 2022

4. Jayanth Reddy Regatti, Aniket Anand Deshmukh, Eren Manavoglu, and Urun Dogan. Consensus clustering with unsupervised representation learning. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. IEEE, 2021
5. Gilles Blanchard, Aniket Anand Deshmukh, Urun Dogan, Gyemin Lee, and Clayton Scott. Domain generalization by marginal transfer learning. *Journal of Machine Learning Research*, 22(2):1–55, 2021
6. Jayanth Reddy Regatti, Aniket Anand Deshmukh, Frank Cheng, Young Hun Jung, Abhishek Gupta, and Urun Dogan. Offline rl with resource constrained online deployment. *arXiv preprint arXiv:2110.03165*, 2021
7. Urun Dogan, Aniket Anand Deshmukh, Marcin Machura, and Christian Igel. Label-similarity curriculum learning. In *European Conference on Computer Vision*, pages 174–190. Springer, 2020
8. Aniket Anand Deshmukh, Srinagesh Sharma, James W Cutler, Mark Moldwin, and Clayton Scott. Simple regret minimization for contextual bandits. In *Exploration in RL, ICML 2019 workshop arXiv preprint arXiv:1810.07371*, 2018
9. Aniket Anand Deshmukh, Urun Dogan, and Clay Scott. Multi-task learning for contextual bandits. In *Advances in Neural Information Processing Systems*, pages 4848–4856, 2017