Debjyoti Paul

November, 2022

CONTACT
dr.debjyotipaul@gmail.com
/in/debjyotipaul385/
www.debjyotipaul.in
www.github.com/debjyoti385

RESEARCH INTERESTS
Spatio-temporal Data Analytics, Database Tuning, Natural Language Processing, Speech Technologies, Representation Learning.

WORK EXPERIENCE

[**March 2020 - *Present***] **Meta, Inc.**, Menlo Park, United States

[Feb 2022 - *Present*] **Senior Research Scientist**

[Mar 2020 - Feb 2022] **Research Scientist**

Scientist in AI Speech team responsible for creating, improving, and deploying state-of-the-art Automated Speech Recognition, Inverse Text Normalization and Punctuation-Capitalization models.

[**Summer 2019**] **Alibaba Group., AI Research Intern**, Sunnyvale, United States

Workload-aware database tuning with Machine Learning.

[**Summer 2018**] **Facebook, Inc., Research Scientist Intern**, Seattle, United States

Improving metrics for Facebook's Search Ranking via rated label inferences with Machine Learning.

[**Summer 2017**] **Amazon AI Research, Research Intern**, New York, United States

Hyperparameter optimization techniques on MXNET.

[**2013-2015**] **Flipkart, Data Engineer**, Bangalore, India

Served as data engineer at data platform team *flipkart.com*, India's biggest e-commerce company.

Facilitated scalable environment for Big data analytics with processing pipelines for batch and stream.

EDUCATION

[**2015-2020**] **Ph.D., University of Utah, School of Computing**, Salt Lake City, Utah, USA

Major : Computer Science & Engineering
Advisor : *Prof. Feifei Li*
Thesis : *Data-Driven Spatio-Temporal Analysis for Multimode Data*
GPA : *3.96/4.0*

[**2011-2013**] **Masters of Technology, Indian Institute of Technology Kanpur**, Kanpur, India

Major : *Computer Science & Engineering*
Advisor : *Late Prof. Sanjeev K Aggarwal*
Thesis : *Multi-constraint Job scheduling in Grid Computing*
GPA : *8.67/10.0 (Rank: 3)*

[**2007-2011**] **Bachelors of Technology**, **West Bengal University of Technology**, Kolkata, India

Major : *Computer Science & Engineering*
College : *Institute of Enginering and Management*
CGPA : *8.93/10 (Rank: <10)*

THESES

[**2020**] **Ph.D. Thesis, Advisor: Dr. Feifei Li**
Data-Driven Spatio-Temporal Analysis for Multimode Data.
The objective of this thesis is to build a framework that enables a data-driven approach for spatio-temporal analysis for multimode data with AI models and facilitate workload-aware support for effective large-scale processing.

[**2013**] **M.Tech Thesis, Advisor: Late Dr. Sanjeev Kumar Aggarwal**
Multi-constraint Job scheduling in Grid Computing
The objective is to efficiently schedule Jobs on Grid Computing to achieve maximum utilization of resources with energy efficient approach. Modeling real world computing & storage grid on Multi-objective Evolutionary Algorithm satisfying hard constraints on jobs constraints, resources constraints, and soft constraints on cost, and energy consumption.

PUBLICATIONS

+ Language Agnostic Data-Driven Inverse Text Normalization, Szu-Jui Chen, **Debjyoti Paul**, Yutong Pang, Peng Su, Xuedong Zhang. *ICASSP 2023. (In Submission).*

+ Improving Data Driven Inverse Text Normalization using Data Augmentation and Machine Translation, **Debjyoti Paul**, Yutong Pang, Sray Chen, Xuedong Zhang, *Interspeech 2022 Show and Tell*, Incheon, Korea, September 2022.

+ Improving Data Driven Inverse Text Normalization using Data Augmentation, Laxmi Pandey, **Debjyoti Paul**, Pooja Chitkara, Yutong Pang, Xuedong Zhang et al. CS arXiv:2207.09674 (2022).

+ Database Workload Characterization with Query Plan Encoders, **Debjyoti Paul**, Jie Cao, Feifei Li, Vivek Srikumar, *48th International Conference on Very Large Data Bases (VLDB 2022)*, Sydney, Australia, 2022. *DOI: 10.14778/3503585.3503600*

+ Semantic Embedding for Regions of Interest, **Debjyoti Paul**, Feifei Li, Jeff M. Phillips, *The International Journal on Very Large Data Bases (VLDBJ 2021). DOI:* 10.1007/s00778-020-00647-0.

+ Bursty Event Detection Throughout Histories, **Debjyoti Paul**, Yanqing Peng, Feifei Li, *35th IEEE International Conference on Data Engineering (ICDE 2019)*, Macau, China, 2019. *DOI:* 10.1109/ICDE.2019.00124

+ Compass: Spatio Temporal Sentiment Analysis of US Election, **Debjyoti Paul**, Feifei Li, Murali Krishna Teja, Xin Yu, Richie Frost, *What twitter says!*, *23rd SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD)*, Aug 13-17, 2017, Halifax, Canada. *DOI:* 10.1145/3097983.3098053.

+ AI Pro: Data Processing Framework for AI Models, **Debjyoti Paul**, Richie Frost, Feifei Li, *35th IEEE International Conference on Data Engineering (ICDE)*, Macau, China, 2019. *DOI:* 10.1109/ICDE.2019.00219.

+ Geotagged US Tweets as Predictors of County-Level Health Outcomes, 2015–2016, Quynh C. Nguyen, Matt McCullough, Hsien-wen Meng, **Debjyoti Paul**, Dapeng Li, *American Journal of Public Health (AJPH), September, 2017*, *DOI:* 10.2105/AJPH.2017.303993.

+ Social media indicators of the food environment and state health outcomes, Quynh C. Nguyen, Hsien-wen Meng, Dapeng. Li, Matt McCullough, **Debjyoti Paul**, Kanokvimankul. P, Nguyen. T, Li. Feifei, *American Public Health Association (APHA), 148, 120-128.*, 2017, *DOI:* 10.1016/j.puhe.2017.03.013.

+ Multi-objective Evolution based Dynamic Job Scheduler in Grid, **Debjyoti Paul**, Sanjeev K. Aggarwal, *The 8th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)*, IEEE, July 2nd - 4th, 2014, Birmingham, UK. *DOI:* 10.1109/CISIS.2014.50.

+ RCached-tree: An Index Structure for Efficiently Answering Popular Queries, Manash Pal, Arnab Bhattacharya, **Debjyoti Paul**, *ACM International Conference on Information and Knowledge Management (CIKM 2013)*, Oct. 27 - Nov. 1, 2013, San Francisco, CA, USA. *DOI:* 10.1145/2505515.2507817.

Position of Responsibilities

[2021] Program Committee Member for 47th International Conference on Very Large Data Bases (VLDB) 2021, Industrial Track.

[2021] Program Committee Member for 37th IEEE International Conference on Data Engineering (ICDE) 2021, Industrial Track.

[2018-2021] Reviewer: IEEE TKDE 2018, IEEE TKDE 2019, IEEE ICDE 2021, PVLDB 2021.

[2016-2019] External Reviewer: EDBT 2016, PVLDB 2019.

[2015-2019] Teaching Mentor and Lecture: Natural Language Processing (CS6340), Data Mining (CS 6140), Database Systems (CS 6530) at the University of Utah. USA.

[2012-2013] Member and organizer of SIGDATA, a Special Interests Group on data management and mining at the IIT Kanpur, India.

[2011-2013] Teaching assistant: Software Engineering (CS455), Fundamentals of Computer Science (ESC101) at IIT Kanpur.

Hackathon Trophies

[2017] HackTheU	HackTheU MLH Hackathon Winner (around 40 teams)
[2016] EMC2 Code	Mars Challenge Hackathon Winner (around 45 teams)
[2015] Goldman Sachs	Air Quality Hackathon 1st Runners-up (around 30 teams)
[2014] InMobi	Freedom Hack Worldwide Hackathon 1st Runners-up (around 160 teams)
[2013] Yahoo	Yahoo HackU 2013 Hackathon Winner (around 40 teams)

Achievements

[2012] Secured **All India Rank 228** out of 160,000+ (99.85 percentile) participants in Graduate Aptitude Test for Engineering (GATE) 2012, Computer Science. Secured **All India Rank 223** out of 150,000+ participants in Graduate Aptitude Test for Engineering (GATE) 2011, Computer Science.

[2011] **All India Rank 7** (99.99 percentile) in Indian Space Research Organization (ISRO) Junior Research Fellow Exam.