Ashutosh Tiwari

669-292-7534 | findashutoshtiwari@gmail.com | Linkedin@ashutosh–tiwari | Homepage/Portfolio | Github@thunderock

Work Experience

Machine Learning Engineer 4 (ML Data Platform) Jul. 2024 – Present
ADOBE FIREFLY *GENERATIVE AI, COMPUTER VISION, LARGE LANGUAGE MODELS*

- Currently part of Machine Learning Data Platform, focused on offline inference for videos and images and feature store.
- Part of a small team that wrote our first version of inference framework.
- Our feature store supports training large machine learning models trained on billions of images and videos.
- Wrote the first data quality framework used to ensure quality of feature generation. This is used across enrichment pipelines to ensure the correctness of output.
- Working on our new Inference Framework, using vLLM and Ray for distributed inference using large language models.

Senior Software Engineer (Data Platform) Sep. 2023 – Jul. 2024
EVOLUTIONIQ *DATA PLATFORM, NATURAL LANGUAGE PROCESSING, REGRESSION*

- Part of Scoring and Quality sub-team, working at the intersection of generative AI, fin-tech, and health sector.
- Responsible for writing data pipelines that ingest, and process data to be ingested, and used for training by our machine learning models.
- Writing pipelines that train our models to predict a claimant's expected time to return to work, ICD extractions from diagnosis notes, alternate VOC recommendations, etc.
- Working on a framework to evaluate that our models are unbiased and fair to different demographic groups.

Career Break for Master's Degree Aug. 2021 – Sep. 2023
Indiana University, Bloomington *Master of Science in Computational Data Science*

- Completed coursework and research in Fairness Aware AI, Network Science, and Natural Language Processing.
- Engaged in projects and independent studies to enhance practical skills in machine learning and data science.

Software Dev Engineer II (ML Platform) Jan. 2019 – Jul. 2021
SWIGGY *DATA SCIENCE PLATFORM, TIME SERIES FORECASTING*

- Was part of team that worked on Feature Store and pipeline which feeds on-demand features to deployed ML models at production scale(4Bn rows, 10K QPS). Pipeline-supported multichannel ingestion, i.e. Spark, Flink and user files etc.
- Founding member of Forecasting and Correlation Platform which was considered by many teams to forecast concerned time series. These forecasts power critical scaling decisions across organizations in real time.
- Led DAQ, a tool used to scrape APIs at scale. Used to collect data for analysis/ model training at a scale of 15 M rows daily.

Software Development Engineer (Search Relevance) Sep. 2017 – Jan. 2019
FLIPKART (a Walmart company) *SEARCH RELEVANCE, QUERY INTENT, NLP*

- Was responsible for improvements/inception of search intent models(CRF/Neural Network based), identifying error classes, coming up with solutions, and fixing them. These models power user search and discovery for millions every day.
- Implemented a FastText based query store classifier, which predicts the category of a tail query.
- Implemented the first workflow to automate training and auto-deployment of various search models in Flipkart. First was written using Luigi and later migrated to Airflow.
- Wrote a generic framework using Airflow which at runtime creates generic dags for different ML models and orchestrates their training to deployment flow, including data and model validations.
- Implemented large scale (4Bn+ datapoints) pipelines using Cascading/HDFS to extract data from user events and then transform it to be used for training these models.

Software Development Engineer Sep. 2016 – Sep. 2017
GROUPON *BACKEND ENGINEERING*

- Worked on a component called Cyclops, an interface between Customer representatives and internal services.
- This service is live in all countries in which Groupon operates.

Software Engineer Sep. 2015 – Aug. 2016
NETSPEED SYSTEMS (Acquired by Intel) *GRAPH ALGORITHMS, NETWORK ON CHIP*

- Led engineering efforts on modules like Polarity based Arbitration, Multi-Cast Filtering, Structural Latency Breakdown, etc.

Publications

Accepted at *NetSci 2023* (Poster Presentation) and *IC2S2 2023* (Parallel Talk) as **first author**

[1] **Ashutosh Tiwari**, Prof. Sadamori Kojaku, Prof. Yong-Yeol Ahn, "Biased Contrastive Learning debiases Graph Neural Networks," *International Conference on Network Science (NetSci)*, 2023. Poster

In this work, we propose a non-parametric contrastive learning framework to learn debiased graph embeddings with respect to sensitive node attributes and structural homophily. Through empirical evaluations on different datasets, we demonstrate that our method offers a better approach to debiasing compared to existing approaches and thus results in more organic recommendations across different GNN architectures.

Education

Indiana University, Bloomington Bloomington, IN
Master of Science in Computational Data Science 3.87/4.0

National Institute of Technology Patna, India
Bachelor of Science in Computer Science 8.32/10.0

Research Experience / Independent Study

- Implementing novel training frameworks that result in unbiased (or fair) recommendation systems from "biased" datasets at IUNI with Prof. YY Ahn and Prof. S Kojaku as part of my Independent Study. Summer 2022 - Summer 2023
- Working as a paid RA for, **"User Intent as a Network"** with Prof. YY Ahn, P Kantak and FB Yara. Project is funded by Kelly Business School. Goal is to be able to quantify and thus act upon that "intent" network. Summer 2022 - Fall 2022
- Was part of NLP Lab@IUB with Prof. D Cavar. Contributed to design of TieML and Events' Timeline modelling using different fine-tuned Large Language Models (LLMs). Fall 2021

Teaching Experience

Teaching Assistant for Network Science (INFO-I 606) with Prof. YY Ahn Spring 2023
Teaching Assistant for Machine Learning (CSCI-B 555) with Prof. R Khardon Fall 2022
Teaching Assistant for Network Science (INFO-I 606) with Prof. YY Ahn Spring 2022

Selected Projects

BiasNet | *Deep Reinforcement Learning, PyTorch, Actor Critic Algorithm, On-policy Model Free* Code/Report
- Learning to fight in Street Fighter II with induced relational bias from differential game scenes.

DeepFoodie | *Python, Tensorflow, Self Supervised Deep Clustering, Deep Learning, Transfer Learning* Code/Report
- Clustering dishes on basis of their ingredient embeddings. These ingredient embeddings are generated by a NN.

BlindNet | *Python, Pytorch, Deep Learning, Transfer Learning* Code/Report
- Image to vector generation on Coco dataset.

Continuous Dominant Set Repair | *C++, Graph Algorithms, Guha and Khuller's Algo., Greedy* Code
- Repairs a broken link in Continuous Dominant Set in $O(\Delta^2)$, where Δ being the avg cardinality of connected graph.

Humana Mays Healthcare Analytics Case Competition | *Boosted Trees, Feature Engineering* Code
- 11th rank on leaderboard. Hosted by TAMU and Humana Mays, 2021.

AnalyticsVidhya Solutions | *Python, Tensorflow, Torch, Time Series, Feature Engineering, Catboost* Code

Investigating Bias Manifolds | *Bias Manifolds, Bias Progression, Measuring Bias, Python, Word2vec* Code/Report

Technical Skills

Search Relevance, Recommendations, Graph Neural Networks, Fairness Aware Modeling, Computational ML, NLP, Computer Vision, DL, Spark, Python, Scala, C++, Contrastive Learning, Large Language Models
Frameworks/Libraries/Tools: Pytorch Geometric, Pytorch Lightening, Pytorch, Tensorflow, Sklearn, Numpy, Pandas, Matplotlib, HDFS, Spark, Kafka, Flink, AWS Sagemaker, DynamoDB, Faiss, vLLM, Ray, Django