Buffalo, NY
+1 (716) 239-3164

Utkarsh Kumar

LinkedIn
utkarshk@buffalo.edu

Research Objective

I am an experienced data scientist with 5+ years experience in industry on projects like setting up data science pipelines and handling massive data ingestion systems.At UB I have credited courses which help me gain the necessary background in distributed systems and efficient algorithm implementations . Throughout all my courses I have thoroughly enjoyed and believed in implementing very efficient algorithms at scale and I am curious as to how can they be extended to AI systems and implementation of DNNs. I am particularly interested in the thrust area of **Near-Memory Acceleration**- to extend present day hardware's capabilities and deep dive into efficient learning algorithms for faster AI implementation.

Education & Relevant Courses

MS, Computer Science, University at Buffalo, NY, CGPA - 3.2 **Sep 2021 –June 2023 (Planned)**

- Distributed Systems, Analysis of Algorithms, Machine Learning, Deep Learning
- Graph mining, String and sequences processing, Simulation of quantum computers (Seminar)

B.E. Electronics and Instrumentation, Jadavpur University, India, CGPA - 3.2 **Sep 2011 — May 2015**

- Digital System Design, Microprocessors, Embedded Systems, Signal Processing, Numerical Methods
- Computer Architecture, Computer Organisation

Experience

Senior Data Scientist **Sep 2019 — Jun 2021**
Global Consumer Banking, CitiBank Bangalore, India

Data Scientist **Sep 2015 — Dec 2017**
MuSigma Business Sol. Pvt. Ltd. Bangalore, India

Skills

Programming Proficiency	C++, Python
Programming Other	SQL, JavaScript
Algorithms	Greedy & Dynamic approaches, Advanced Graph algorithms, Advanced string and stream algorithms
Data Science & Learning Systems	Data Wrangling & ETL, Data Mining, Deep Learning, Machine Learning

Projects

Motif based sub-graph mining:

- Studied the application of motifs in fair graph mining
- Analyzed subgraphs based on different types of graph motifs. Studied the data mined by motif decomposition and its fairness in discovering protected variables such as discovering groups of females in a Facebook friendship network

Basic graph algorithms:

- Implemented BFS and its optimized variant to study speedups
- Implemented graph transformations and edge based segmentation algorithms
- Studied variants of graph traversals and suggested efficient way of calculating all BFS traversals in a graph
- Implemented and compared between PageRank and HIITs

Pub-Sub distributed system over Flask and Mongo DB

- Created a pub-sub system using flask framework and mongo db and dockerized the application
- Application had 3 brokers handling 2 Mongo servers to fetch jokes from external apis when publishers request for it, and then queue the newly published data for subscribers to read. The application was scaled and demoed for 5 publisher accounts and 5 subscriber accounts

String Algorithms: Experience in implementing systems which can search over massive amounts of text or sequence data in provable linear times

- Efficiently implemented Boyer-Moore heuristic for linear time search speed ups
- Implemented suffix tree for search over massive collection of multiple patterns. The implementation consisted of failure links to enable speedups in search
- Implemented and used GSA and LCP arrays to fast string search
- Given compressed string stats, implemented fast BWT (FM Index) algorithm to search over massive texts
- Implemented fast hashing based string containment algorithm to search one string within another

Graph community detection based on machine learning algorithm

- Studied machine learning based approaches for detection of communities in graph
- Suggested improvements based on graph characteristics such as node centrality measures

Established ETL pipelines (Industry experience)

- Established pipelines to ingest unstructured customer conversation data and stage them into structured tables
- Fed these tables into machine learning models in order to quantify sentiments. The pipeline had to have various pre-processing and anomaly checks to maintain output quality
- Developed front end dashboards which derived meaningful insights from the model output and fed to business consumers in real time