Arpita Saha

arpitasaha1996@gmail.com | 9374776531 | LinkedIn | Github | Scholar | Website

EXPERIENCE

Research Associate, Brandeis University, USA. ***October 2023-Present***

- Hypothesized the problem of automatic knob tuning in NoSQL databases (LSM engine) for dynamic workload optimized database design and robust configuration selection. Designed a tentative Machine Learning solution and submitted a grant proposal to Amazon Research Awards (current project)
- Designed and presented a poster entitled: Toward Workload-Aware Self-Designing LSM-Engines at NEDB Day 2024.
- Working in a project that aims to achieve LSM memory profiling for different data structure implementations of the memtable. Studying and implementing memtable data structures in CASSANDRA and ROCKSDB
- Paper entitled "KVBench: A Key-Value Benchmarking Suite" accepted for publication in DBTest 2024. This paper introduces a workload generator tool used to stress test NoSQL data systems.: https://dl.acm.org/doi/10.1145/3662165.3662765

Graduate Research Assistant, Ohio State University, USA. *January-August 2023*

- Piloted the project for Covid-19 Mortality Prediction and Patient Phenotyping from large-scale EHR data.
- **Published first-authored paper** entitled "A Multi-Layered GRU Model for COVID-19 Patient Representation and Phenotyping from Large-Scale EHR Data" accepted at ACM-BCB 2023 (acceptance rate 29%): link
- Developed a GRU-based time-series **deep learning model** (only **11k** parameters) to predict COVID-19 patient mortality outcome with an ROC AUC of **97%** that outperforms all baselines (having around 700k parameters).
- Uncovered 4 distinct phenotypes by **clustering** strong patient representation embeddings and analyzed trends across phenotypes to identify risk factors related to mortality for **efficient resource allocation** during pandemic.
- Built an **interactive desktop application** to visualize time-series patient data using PyQT5 python module.

Graduate Teaching Assistant, Ohio State University, USA. *January-December 2022*

- **Communicated effectively** and built good rapport with instructor and students, enabling smooth class conduction.
- Contributed to the development of exam materials for appropriate and timely student evaluation.
- **Mentored and managed** a class of 100 students, and maintained their course roster and grade sheets.

Undergraduate Researcher, Bangladesh University of Engineering and Tech, BD. *April 2019 - May 2021*

- Designed a novel semi-supervised variational auto-encoder deep learning model to impute missing taxa into gene trees as a member of a 3-person team, published the work as a co-author in RECOMB 2022: link.
- Used NLP techniques such as masked language modeling and positional encoding to improve performance.
- Utilized Python (Numpy, Pandas,Tensorflow) to code an end-to-end analysis pipeline and conduct all experiments.

EDUCATION

The Ohio State University,USA - *MS in Computer Science and Engineering (CGPA: 3.81/4)* *August 2023*

- Thesis research on Machine Learning and AI in HealthInformatics: link
- Courses: Algorithms, Operating Systems, Programming Languages, Computer Vision, NLP, etc.

Bangladesh University of Engineering and Technology, BD - *BS in CSE (CGPA: 3.82/4)* *February 2021*

- Courses: Object-Oriented Programming, Software Engineering, Networking, Database, Machine Learning, etc.

SKILLS & INTERESTS

- Languages: Python, Java Script, C, C++, Java, MATLAB, SQL, HTML, CSS, SHELL
- Frameworks/Libraries: Django, Java Swing, PyQT5, MySQL, SQLite, PyTorch, Pandas, Matplotlib, Express
- Tools/Infrastructure: Git, SLURM, Linux, UNIX, Java Unit Testing, Agile, Scrum.

PROJECTS

Studying the effect of Sparsification and Quantization on Large Language Models *June 2024*

- Sparsified TinyLlama-1.1B with sparse-gpt and quantized to 8 bits: link
- Studied the effect on accuracy of token predictions for datasets such as hellaswag, arch_challenge, mmlu, gsm8k, TruthfulQA, Winogrande

RESTful API for data exchange about Products and Order *June 2024*

- Built a RESTful API using Node.js and Express: link
- Stateless data exchange in JSON in a client-server architecture about products and orders
- GET, POST, PUT, DELETE, PATCH endpoints supported.

O-H-I-O Pose Detection from Live Video Input (Computer Vision) *December 2022*

- Built a desktop app using Python for collecting images to curate a dataset by collecting live video feed using webcam at different lighting and background conditions.

- Leveraged the frames from videos to build MEI, MHI images and calculate similitude moments, which were used as features for the KNN classifier that detects the correct pose.

Rating Software for Alpha Credit Rating Company *March 2021*

- Built a software to calculate transition probability from one rating to another in a year based on past data of companies using Java Swing and MySQL.
- Planned and executed **full-stack** development of the Software, including relational database design.

Tour Planner Website (Software Engineering Project) *January 2019*

- Built a website using Django and SQLite for planning a tour given destination and time budget: link.
- Used Traveling Salesman Problem as backend algorithm; incorporated search and admin privileges.

PERSONAL ACHIEVEMENTS

Awardee, Anita B. Org Scholarship for Attending GHC 2022 *January 2022*

Vice President, Computer Society, IEEE BUET SB: *July 2019 - February 2021*

Co-founded the organization and organized technical workshops and coding contests to improve interaction among computer science enthusiasts across various universities of the country.

Awardee, BUET Dean's List Award and Merit Scholarship for top 10 in Computer Science *2016, 2018*