Md Kamruzzaman (Methun)

mhmethun@gmail.com | (509) 336-9938 | Folsom, CA, USA | linkedin.com/in/mhmethun

Summary

Ph.D.-trained researcher and software developer with 5+ years in applied machine learning, solving cross-domain problems including infectious disease, health anomaly detection, and drug discovery in science and technology. Expert in SFT, PEFT, and generative AI (VAE, WGAN, Diffusion), achieving $4x$ performance boosts and 95% efficiency gains using GPU clusters. Proficient in Python, C++, SQL, and MPI/OpenMP. Led 6-person teams, collaborated with Google, and seeks to advance ML innovation.

Skills

- Machine Learning: SFT, PEFT, LoRA, VAE, WGAN, Diffusion Models, Ensemble Models, Hierarchical Models
- Data Science: ETL, Topological Data Analysis (TDA), Statistical Modeling, Graph-Based Algorithms
- Programming: Python, C++, CUDA, NodeJS, D3.js, Java, HTML/CSS, JavaScript, SQL
- Tools & Frameworks: Huggingface (dnaBERT2, EVO2, UTRLM, molgpt), PyTorch, TensorFlow, HPC Clusters (SLURM)
- Additional: Privacy-Preserving ML, Data Visualization, Full-Stack Development, Git CI/CD, MLFlow, Docker
- Soft Skills: Collaborative, Team Player, Team Leadership, Bias for Action, Results-Driven

Professional Experience

Postdoctoral Researcher (Applied Machine Learning), Sandia National Laboratories, Livermore, CA — May 2023–Present

- Advanced multilayered semi-realtime anomaly detection research on health sensor data (collected from Fitbit and Apple Watch) utilizing GPU and generative AI
  - Reduced cokurtosis tensor compute time by $5x$ on wearable health data (steps, heart-rate, sleep, etc) using CUDA on NVIDIA H100 GPU.
  - Improved anomaly detection accuracy for a $45,000$-population (digital twins) by advancing generative AI models (VAE, WGAN, Diffusion), minimizing model uncertainty.
- Spearheaded molecular discovery research for a specific group of chemical compounds by fine-tuning LLM
  - Enhanced lab efficiency 95% by fine-tining Huggingface/molgpt with $1,000$ candidate SMILEs strings, and generated $10k$ new molecules.
  - Discovered two industry-manufacturable candidate molecules utilizing an ensemble-based predictive and rule-based recommendation model from $10k$ new molecules.
  - Collaborated with cross-domain researchers to decipher actionable requirements and model explainability using SHAP.
- Led bioinformatics research on drug-repurposing and $5'$UTR mRNA sequence classification
  - Tool development for FDA-approved drug candidates, improving drug repurposing strategies, using human-virus PPI, transcriptomic, and viral gene-drug interaction data.
  - Accelerated $5'$UTR mRNA sequence classification by $4x$, reducing training time, using SFT and PEFT with LoRA across $10M$ samples on a 10-node $A100$ GPU cluster.
- Enhanced lattice QCD matrix element modeling accuracy at short distances and large momentum using generative AI techniques.
- Mentor undergraduate and graduate interns and work closely to expedite the project delivery.

Postdoctoral Researcher (Applied Machine Learning), University of Virginia, Charlottesville, VA — Oct 2020–May 2023

- Led cross-domain research on machine learning applications in healthcare, implementing novel algorithms for detecting infection sources and risk assessment on IRB-approved HIPAA-compliant MRSA data.
  - Streamlined data preparation for model training by utilizing an ETL process to integrate daily data from two hospitals into a common data model.
  - Constructed a patient-provider temporal contact network, extracted latent features that improved model AUC by 15%.
  - Enhanced model accuracy by 7% on unbalanced EHR datasets by designing a novel hierarchical model algorithm.
- Strengthened healthcare application performance by implementing privacy-preserving ML solutions in collaboration with the Google research team.
- Identified infection sources and missing cases using an agent-based model on an HPC cluster (MPICH) with SLURM, leveraging claim and prescription data.
- Mentored undergraduate and graduate students and interns.

Research Assistant (Data Science), Washington State University, Pullman, WA — Aug 2015–Sep 2020

- Led research in algorithm development to extract actionable insights from complex and high-dimensional plant phenomics and hospital patient data.
  - Developed Hyppo-X, a CPP-based data analysis framework with integrated Topological Data Analysis (TDA) algorithms and interactive visualization tools, uncovering complex patterns in high-dimensional datasets.
  - Enabled breakthroughs in maize phenomics research by engineering a graph-based data analysis algorithm to extract scientific hypotheses from structural properties.
  - Delivered actionable insights for agricultural research by implementing statistical modeling for genotypic analysis under environmental stress.
  - Enhanced clinical insights by integrating NodeJS and D3.js for advanced data visualization of hospital patient trajectories.
  - Taught undergraduate students VeriLog, and software engineering.

Software Development Engineer, KBGroupUK & NRG, Dhaka, Bangladesh — May 2009–Jul 2015

- Built full-stack e-commerce websites using .NET, HTML/CSS, and JavaScript, increasing ticket sales by 13%.
- Led a 6-person team from 2013 to 2015, driving 5% year-over-year revenue growth.

Internship Experience

Applied Scientist Intern, Amazon, Seattle, WA — Jun 2019–Sep 2019

- Optimized ML models (ARIMA, LSTM) for sales predictions, improving accuracy by 3% utilizing the Amazon SageMaker.

Research Intern, Pacific Northwest National Laboratories, Richland, WA — Jul 2018–Aug 2018

- Engineered C++ solutions with SPDLOG, resolving Python asynchronous bottlenecks.

Education

Ph.D., Computer Science, Washington State University, Pullman, WA — Aug 2015–Aug 2020
B.Sc., Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh — Mar 2004–Mar 2009

Additional Information

- Proficient in collaborating across interdisciplinary teams and presenting complex technical results to diverse stakeholders.
- Available for relocation and authorized to work in the United States without VISA sponsorship (GC holder).