# Methun Kamruzzaman

mhmethun@gmail.com | github.com/xperthut | 509-336-9938 | linkedin/in/mhmethun

## SKILLS

- `Object oriented:` C++, Python, Javascript, C#
- `Script:` Python, R, Javascript
- `Web:` nodeJS, electronJS
- `Database:` PostgreSQL, MSSql, MySql
- `Visualization:` D3, iGraph
- `IDE:` Xcode, VS Code, R Studio, Jupyter notebook, Netbeans
- `OSs:` Windows, MacOSX, Linux
- `Tools:` MS Office, xFig, GnuPlot

## EXTRACURRICULAR ACTIVITIES

- Volunteered in research on COVID-19 to generate narrative using subgraph mining (Aug–Sep' 2020).
- Web content management of ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB'2020).
- Volunteered to develop a "Feed The Hungry" programs associated with Community Action Center in Pullman WA.
- Participated regularly in environmental clean up and beautification projects around Pullman and with various interfaith groups.
- Volunteered in assisting learning disabled students in usage and computer skills.

## ACADEMIC PROJECT EXPERIENCE

- Utilized machine learning methods to detect highly productive soybean varieties (using `svm`, `random forest` model).
- Spectrum analysis to detect contagion node in large disease epidemic network (using `R`).
- Developed a software to map more than 900k pear/peach reads to the (Cherry) reference genome (more than 300k long).
- Detected cancer genes from gene expression dataset using Persistent homology of the Mapper framework.

## EDUCATION

### Ph.D. | Computer Science

Washington State University | GPA: 3.82 | 2020

- Advisors: Ananth Kalyanaraman (CS), Bala Krishnamoorthy (Math)
- Dissertation: *Topological Data Analysis for Computational Phenomics: Algorithms and Applications (Semi-automated hypotheses extraction using unsupervised learning)*

### B.Sc. | Computer Science & Engineering

Bangladesh University of Engineering & Technology, Bangladesh | GPA: 3.66 | 2009

## EXPERIENCE

### Biocomplexity in UVA | Postdoctoral researcher | October, 2020-Current

- Deep factorization based recommendation system predicts more effective hybrids across farming locations and years.
- ML based modeling to analyze risk factors for MRSA infected patients (from EMR data), that helps to reduce the number of testings.
- Agent based modeling helps to identify the source of infection and missing infections, that suggest precautions of hospital acquired infections.
- Benchmark analysis of Privacy-preserved distributed ML models helped to launch *Google Health Studies app*.

### Washington State University | Research assistant | Fall 2015 - Summer 2020

- Developed a Data Exploratory framework in `C++` (named `Hyppo-X`) based on Topological Data Analysis (`TDA`) with interactive visualization capability (using `electronJS` and `D3`) that reveals the high order structure of high dimensional complex dataset.
- Developed theories and algorithms based on the structural properties (flares and paths) of the graph that help domain scientists to extract plausible hypotheses.
- My framework discovered new patterns in maize phenomics data (sensitivity to location (KS/NE)) and also in patient trajectories in hospitals.
- My subgraph mining method helps domain scientist to generate narrative.
- Statistical modeling to reveal the genotypic diversity on environmental stress.

### Amazon Inc. | Machine Learning intern | Summer 2019

- My exploratory data analysis (using `SQL` and `R`) has revealed more than 10 issues on database and 4 of them were considered during data lake design.
- Designed forecasting model (using `ARIMA`, `LSTM`) on sales and customer data using `keras` and `scikit-learn` library.
- My forecasting model produced 3% more accuracy compared to existing model.
- Training on leadership principles in order to take ownership of the project.

### Pacific Northwest National Laboratory | Research intern | Summer 2018

- Developed a prototype in `C++` for benchmark analysis of an existing open source framework named SHAD which runs in a distributed cluster.
- For benchmark analysis, I used another open source software named SPDLOG and embed it with SHAD.
- My analysis (using `Python Pandas`) identified a performance issue of an asynchronous function call (process-2 calls the method 5× more compared to process-1) when running the framework for triangle count in a dense graph with trillion nodes.

### NRG+KBGroup | Software Development Engineer | 2009 – 2015

- iOS app embedded with localized and optimized marketing content, to markets globally. Cost efficiency as well as the use of creativity and reduced time to market.
- Developed one of the `E-Commerce` websites from start to finish (`ASP`, `HTML`, `CSS`, `JS`). New website increased in ticket sales by more than 13% .
- Led a 6 person development team (2013-15). Company revenue increased 5% YoY.

## PUBLICATIONS

- Privacy-first health research with federated learning. [medrxiv: *https://doi.org/10.1101/2020.12.22.20245407*]
- `Hyppo-X`: A Scalable Exploratory Framework for Analyzing Complex Phenomics Data. [TCBB'19: *https://ieeexplore.ieee.org/document/8880514*]
- Interesting Paths in the Mapper Complex. [JOCG'19: *https://doi.org/10.20382/jocg.v10i1a17*]
- *https://scholar.google.com/citations?user=SVj4svcAAAAJ&hl=en*