EFFAT FARHANA

effat.farhana@vanderbilt.edu | Website | LinkedIn | Google Scholar

RESEARCH VISION

My research interests include *Explainable Artificial Intelligence (XAI)*, *Data Mining*, *Applied Natural Language Processing (NLP)*, and *Machine Learning*. My long-term goal as a researcher is to bridge **human-computer interaction** with **machine learning (ML) and artificial intelligence (AI)**. To that end, my research focuses on education and healthcare domains. Specifically, my work focuses on mining educational software (e.g., MOOCs, online learning platforms) to derive **data-driven heuristics** and designing **interpretable machine learning algorithms**.

PROFESSIONAL EXPERIENCE

Postdoctoral Research Scholar — Jul 2021 - Present
Vanderbilt University — Nashville, TN

Graduate Research Assistant (RA) — Aug 2020 - May 2021
North Carolina State University — Raleigh, NC

Graduate Teaching Assistant (TA) — Aug 2016 -Aug 2020
North Carolina State University — Raleigh, NC

Graduate Research Assistant (RA) — Aug 2015- Aug 2016
North Carolina State University — Raleigh, NC

Lecturer, Computer Science — Apr 2011 - Dec 2014
Ahsanullah University of Science and Technology — Dhaka, Bangladesh

EDUCATION

North Carolina State University, Raleigh, NC — Aug 2015 - April 2021
Doctor of Philosophy (Ph.D.) in Computer Science
Advisor: Dr. Collin F. Lynch
Dissertation title: Science Reading Behavior of Middle School Students within a Digital Literacy Platform

Bangladesh University of Engineering & Technology, Dhaka, Bangladesh — Feb 2011
Bachelor of Science (B.Sc.) in Computer Science and Engineering

AWARDS

- Selected as a **Rising Star in Data Science- Jan 2021** organised by Center for Data and Computing (CDAC), University of Chicago.
- Travel award to attend Women in Machine Learning (WiML) workshop, co-located with NeurIPS 2020 (virtual).
- Travel award for CRA-WP Widening Participation Early and Mid Career Mentoring Workshop, 2020.
- Awarded scholarship by Women in Computer Science (WiCS), NCSU to attend Grace Hopper Conference, 2018.
- Awarded scholarship by ACM Richard Tapia Celebration of Diversity in Computing, 2020 to present poster in ACM Student Research Competition.

PUBLICATIONS

Peer-Reviewed Journals and Conference Publications

1. Predictive Student Modelling in an Online Reading Platform
   Effat Farhana, Teomara Rutherford, and Collin F. Lynch.
   Twelfth AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-2022, Collocated with AAAI) (In Press).
2. Understanding Reading Behaviors of Middle School Students
   Effat Farhana, Teomara Rutherford, and Collin F. Lynch.
   Proceedings of the Seventh ACM Conference on Learning @ Scale (L@S 2020).
3. Associations Between Self-Regulated Learning Strategies and Science Assignment Score in a Digital Literacy Platform
   Effat Farhana, Teomara Rutherford, and Collin F. Lynch.
   Proceedings of the International Conference of the Learning Sciences (ICLS 2020).
4. Investigating Relations between Self-Regulated Reading Behaviors and Science Question Difficulty
   Effat Farhana, Teomara Rutherford, and Collin F. Lynch.
   Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020).

5. Gang of eight: A Defect Taxonomy for Infrastructure as Code Scripts
   Akond Rahman, **Effat Farhana**, Chris Parnin, and Laurie Williams.
   Proceedings of the 42nd International Conference on Software Engineering, (ICSE 2020)
6. The 'as code' Activities: Development Anti-patterns for Infrastructure as Code
   Akond Rahman, **Effat Farhana** and Laurie Williams.
   Empirical Software Engineering. 25, 3430–3467 (EMSE 2020).
7. Synthesizing Program Execution Time Discrepancies in Julia Used for Scientific Software
   Effat Farhana, Nasif Imtiaz and Akond Rahman,
   IEEE International Conference on Software Maintenance and Evolution (ICSME 2019)
8. Challenges with Responding to Static Analysis Tool Alerts
   Nasif Imtiaz, Akond Rahman, **Effat Farhana** and L. Williams.
   IEEE/ACM 16th International Conference on Mining Software Repositories (MSR 2019)
9. Snakes in Paradise?: Insecure Python-related Coding Practices in Stack Overflow
   Akond Rahman, **Effat Farhana**, Nasif Imtiaz.
   IEEE/ACM 16th International Conference on Mining Software Repositories (MSR 2019)
10. Biogeography-based Rule Mining for Classification
    Effat Farhana and Steffen Heber.
    Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2017).
11. Constrained sequence analysis algorithms in computational biology
    Effat Farhana, and M. Sohel Rahman.
    Information Sciences 295 (2015).
12. Doubly-constrained LCS and hybrid-constrained LCS problems revisited
    Effat Farhana, and M. Sohel Rahman.
    Information Processing Letters 112.13 (2012).
13. Finite Automata Based Algorithms for the Generalized Constrained Longest Common Subsequence Problems
    Effat Farhana, Jannatul Ferdous, Tanaeem M. Moosa, M. Sohel Rahman.
    17th International Symposium of String Processing and Information Retrieval, (SPIRE 2010).

Peer-Reviewed Workshop Publications

1. Predicting Post-College STEM Enrollment from Middle School Clickstream Data
   Effat Farhana, Maaz Saleem Kapadia, Wenjia Cao, and Collin F. Lynch.
   Workshop on Scientific Findings from the ASSISTments Longitudinal Data Competition: Educational Data Mining (EDM 2018).
2. A Parallel Island Model for Biogeography-based Classification Rule Mining in Julia
   Samuel Ebert, **Effat Farhana**, and Steffen Heber.
   Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO 2018).

Peer-Reviewed Poster

1. Feedback and Self-regulated Learning in Science Reading
   Effat Farhana, Andrew Potter, Teomara Rutherford, and Collin F. Lynch.
   Proceedings of the 14th International Conference on Educational Data Mining (EDM 2021).

Doctoral Consortium

1. Self-Regulated Learning and Science Reading of Middle-School Students
   Effat Farhana, Teomara Rutherford, and Collin F. Lynch.
   Doctoral Consortium. The 13th International Conference on Educational Data Mining (EDM 2020).

SELECTED RESEARCH EXPERIENCE

Vanderbilt University

- Working in a cross-disciplinary research project between departments of computer science, psychology, and medical school at Vanderbilt. Analyzing log data in a game-based learning environment for kids on the autism spectrum (ASD) and connect their in-game performance with social-skills.

North Carolina State University

- **Developing and Applying Data Mining and ML Methods in Educational Applications**
  - Developed an interpretable ML model to capture students' sequential activities and predict their per-

formance on questions (EAAI/AAAI 2022) (In Press).
- Applied sequential pattern mining technique to understand reading patterns for 12.5K science and 16.2K social study student data (L@S 2020).
- Applied data-driven methodologies to understand how students adapt their reading behavior (i) with varied question difficulty (EDM 2020), (ii) receive feedback from teacher (EDM 2021).
- Predicted students' college enrollment and first job after college from students' middle school interaction within a mathematics educational software (EDM 2018 Workshop).

- **Interpretable ML Algorithm for Classification**
  - Designed and implemented a rule mining algorithm for the classification task utilizing an evolutionary algorithm, Biogeography-based Optimization (BBO). The goal was to optimize the trade-off between accuracy and interpretability in ML algorithms. My proposed algorithm performed better compared to baseline classification algorithms (GECCO 2017).
  - Implemented a parallel version of the proposed algorithm utilizing Julia programming language's parallelization features (GECCO 2018).
- **Software Engineering**
  - Mined Stack Overflow posts to understand performance issues in Julia programming language (ICSME 2019).
  - Conducted an empirical analysis to derive bug categories in configuration management scripts (ICSE 2020).

TEACHING

North Carolina State University

- Teaching Assistant (TA)
  -TA for graduate level courses: Design and Analysis of Algorithm ($\sim$ 200 students), Database Management Concepts and Systems ($\sim$ 150 students), Software Engineering ($\sim$ 30 students), Artificial Intelligence ($\sim$ 60 students), and undergraduate level Data Structure and Algorithm ($\sim$ 150 students) course.
  - Created assignments and exam questions, held office hours, and graded.

Ahsanullah University of Science and Technology

- Lecturer in Computer Science and Engineering
  - Instructor for undergraduate level introductory programming language, design and analysis of algorithm, and compiler courses.

MENTORING

Undergraduate Research Mentoring **Aug 2017- Aug 2018**
North Carolina State University *Raleigh, NC*

- Co-advised Samuel Ebert on interpretable ML algorithm project. This work resulted in a student's lead author paper (GECCO 2018).

MEMBERSHIP

- ACM Professional Member Membership No.: **4570793**
- AAAI Member

PROFESSIONAL SERVICE AND ORGANIZATIONAL EXPERIENCE

- Sub-reviewer at EDM 2021
- Shadow PC at Mining Software Repository (MSR), 2021 Conference
- Reviewer at ICLS 2020
- Judge for VandyHacks 2021 (Vanderbilt's premier student hackathon)
- Ph.D. panel member at Doctoral Recruiting Day 2020, NCSU
- Volunteered at NC State International Graduate Student Orientation 2017