A S M Shahadat Hossain

2900 Old 63 S Apt. 10-301, Columbia, MO 65201, USA
asm-shahadat.github.io linkedin.com/in/asm-shahadat
asmhossain20@gmail.com +1 (779) 902-0043

Education

University of Missouri — Columbia, MO, USA
Ph.D. (ongoing) in Computer Science — *Jan 2025 - Present*

Northern Illinois University — DeKalb, IL, USA
M.S. in Computer Science — *Aug 2024*
Thesis: Evaluating Computational Reproducibility of Jupyter Notebooks Using ML and NLP.

Rajshahi University of Engineering & Technology — Rajshahi, Bangladesh
B.Sc. in Computer Science & Engineering — *Feb 2018*
Thesis: Fake News Detection within a Static Dataset using Supervised Machine Learning Algorithms.

Research Interests

My research interests include but are not limited to:

- **Data Management**: Computational Reproducibility, Data Engineering, Data Management Systems.
- **Machine Learning & Natural Language Processing**: Applied Machine Learning, Large Language Models.
- **Software Engineering**: Program Analysis, Applications of Artificial Intelligence in Software Engineering.

Publications

- **Hossain, ASM Shahadat**. "Customer Segmentation using Centroid Based and Density Based Clustering Algorithms." In *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)*, pp. 1-6. IEEE, 2017. [**Cited by: 49**]

  Work in progress:
  "Building Novel Methods to Evaluate Computational Reproducibility of Jupyter Notebooks."

Work Experience

Research Assistant (Computer Science) — Jan 2025 - Present
University of Missouri — *Columbia, MO, USA*

Working in an NSF (National Science Foundation) funded research project:

- Designing and Conducting Computational Experiments.
- Collaborating with cross-functional teams of scientists from the University of Illinois at Urbana-Champaign and the University of Notre Dame.
- Preparing Scientific Presentations and Research Paper for Conferences.
- Assisting in setting up Research Computing Lab.

Teaching Assistant (Computer Science) — Aug 2020 - May 2024
Northern Illinois University — *DeKalb, IL, USA*

Held office hours, graded weekly assignments, quizzes, final exams, and proctored exams for courses:

- CSCI 330: UNIX and Network Programming; Fall 2021, Fall 2022, Spring 2023, Fall 2023, Spring 2024.
- CSCI 463: Computer Architecture and Systems Organization; Spring 2022.
- CSCI 240: Computer Programming in C++; Fall 2020, Spring 2021.

Graduate Staff Assistant (Data Science & Business Intelligence) Jun 2022 - Sep 2022
Northern Illinois University *DeKalb, IL, USA*

- Generated occupations and skills-based suggestions for graduate students by applying Data Science on Labor Market Data supplied by Emsi Burning Glass and 18K course syllabus documents.
- Ran processes on a Hybrid CPU/GPU Computing Cluster having 60 nodes capable of 30 teraFLOPS.
- Detected and Corrected errors in 13 Python scripts caused due to System Migration (Windows to Linux).
- Performed Data Migration from flat files to Snowflake (Cloud Computing-based Data Warehouse).
- Generated Data Visualization using Microsoft Power BI for Graduate Career and Professional Development portal.

Specialist (Customer Insights Platform & Operational Support Systems) May 2019 - Feb 2021
Robi Axiata Limited *Dhaka, Bangladesh*

- Developed at least 10 scripts using Python, PL/SQL, Linux Shell, Windows PowerShell to automate manual jobs to save at least 4 man-hours per day that is approximately $8K per year.
- Managed Oracle, MySQL, Microsoft SQL Server Database Systems.
- Provided Server Administration Assistance (Linux: Red Hat Linux, SUSE Linux, and Windows: 2012 R2, 2016).
- Initiated Telecom Network Change Request, Monitored and Ensured Necessary Follow-up.
- Extended operational support in Agile and Scrum projects where more than 100 daily manual tasks were automated with a budget of more than $1 million.
- Analyzed network performance data of approximately 50 million subscribers by using tools and scripts.
- Performed System Administration of Telecommunication (Transmission, Radio, and Core) Network Management Systems (NMS) and Customer Insights Platforms - 4 developed by Nokia, 6 developed by Huawei, and 2 developed by Ericsson.

Jr. Engineer (Business Intelligence & Data Warehouse) Aug 2018 - May 2019
Grameenphone IT (Wipro Bangladesh) Limited *Dhaka, Bangladesh*

- Performed Developments and Operations of daily ETL (Extract, Transform, Load).
- Investigated Data Trend Mismatch from a revenue of approximately $40K per day.
- Generated monthly High-Value customers' database from 83 million subscribers using PL/SQL.
- Integrated new processes in Oracle Data Integrator (ODI) at least 2 times a month.
- Monitored daily Operating System and Database performance of a cluster of 8 Linux servers.

Project Engineer (Transmission Optimization & Modernization) Apr 2018 - Jun 2018
Robi Axiata Limited *Dhaka, Bangladesh*

- Ensured daily implementation support in Project (Agile) Management and Vendor Management.
- Initiated and monitored at least 10 Network Change Requests initiation and execution per day for 3G to 4G network upgradation.

Projects

- **Evaluating Computational Reproducibility of Jupyter Notebooks Using ML and NLP.**

  Several methods were developed to compare cell outputs between original and rerun Jupyter Notebooks and evaluate their reproducibility. Natural Language Processing was used to extract textual features from the markdown texts and Machine Learning models were built to classify the notebooks automatically based on their reproducibility.

- **Reproducibility of Visualizations in Jupyter Notebooks: Algorithms vs Programmers.**

  Changes in visualizations upon rerunning Jupyter Notebooks were tracked and categorized. A set of survey questions was proposed to understand the differences between programmers' perceptions and algorithms for evaluating the reproducibility of image outputs.

- **'Reproducibility' in Computer Science Conferences.**

  Texts from 55 Computer Science conference websites mentioning reproducibility and 250 research papers accepted in such conferences were analyzed to understand how computer science conferences assess the reproducibility of the works submitted.

- **Weather Station using Raspberry Pi and Environmental Sensors.**

  In this Software Engineering project, an IOT-based Web Interface to report temperature, humidity, and air pressure using Raspberry Pi was developed. To make the application elegant, several good software engineering practices were implemented to ensure reliability, scalability, and robustness.

- **Clinical Trials Categorization using ML and NLP based on their Media Attention.**

  15 classification models including general purpose classifiers such as Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), Naive Bayes (NB) as well as ensemble methods such as Random Forest (RF), Bagging, Adaboost, and XGBoost were used to classify clinical trials based on their Altmetric Attention Scores (AAS). The highest test accuracy was found to be 94.5% for the numerical features and 80.1% for the text features by using the XGBoost (Extreme Gradient Boosting) classifier.

- **Customer Segmentation using Centroid Based and Density Based Clustering Algorithms.**

  This project illustrated the idea of applying density-based clustering algorithms such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise) for customer segmentation besides using centroid-based algorithms such as K-means. Clusters were formed among 440 customers based on their spending habits to identify the high-value customers.

- **Fake News Detection within a Static Dataset using Supervised Machine Learning Algorithms.**

  Machine Learning & Natural Language Processing (ML-NLP) based models were built to detect fake news within a static dataset. 3 Machine Learning Classifiers - Support Vector Machine (SVM), Logistic Regression (LR), and Stochastic Gradient Descent (SGD) were used. Among them, the highest accuracy was observed as 91.1% for Logistic Regression with a precision of 87.1% and a recall of 95.6%.

Skills

- **Data Management**: Enterprise Data Warehousing, Business Intelligence Data Integration, ETL (Extract, Transform, Load) in Oracle, MySQL, and Microsoft SQL Servers.
- **Data Analysis**: Machine Learning and Natural Language Processing using Python libraries such as NumPy, Pandas, Scikit-learn, NLTK, Gensim, Matplotlib, Seaborn, OpenCV, SpaCy, SciPy, TensorFlow, PyTorch, and Keras.
- **Data Visualization**: Tableau, Microsoft Power BI, Microsoft Excel, and JavaScript D3.
- **Automation**: Tasks automation using Python, PL/SQL, JAVA, Bash, Windows PowerShell, and PHP.
- **System and Server Administration**: Red Hat Linux, SUSE Linux, and Windows.
- **Interpersonal Skills**: Leadership, teamwork, conflict management, patience, and communication.

Service

- **Student Program Committee**:
  - 13th International Conference on the Theory of Information Retrieval with SIGIR (ICTIR), 2023.
- **Reviewer**:
  - The 5th International Conference on Electrical, Computer and Energy Technologies (ICECET), 2025.
  - The 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2024.
  - The 4th International Conference on Electrical, Computer and Energy Technologies (ICECET), 2024.
  - The 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2023.
  - The 3rd International Conference on Electrical, Computer and Energy Technologies (ICECET), 2023.
  - International Conference on Artificial Intelligence, Control, Data Sciences and Applications (ACDSA), 2023.
- **Judge**:
  - Poster Presentation at Show Me Research Week, University of Missouri, 2025.
  - Poster Presentation at Missouri Junior Science and Humanities Symposium (JSHS), University of Missouri, 2025.
  - Conference on Undergraduate Research and Engagement (CURE), Northern Illinois University, 2024.

- **Problem Setter and Judge**, RUET Analytical Programming Lab (RAPL), 2015-2017.
- Other services:
  - Volunteer (Evaluation), STEM Fest sponsored by Meta and hosted by Northern Illinois University, 2023.
  - Volunteer (Evaluation), STEM Fest sponsored by Meta and hosted by Northern Illinois University, 2022.
  - Volunteer (MobilePack), Feed My Starving Children, 2021.
  - Volunteer (Laser Fountains & Holograms), STEM Fest sponsored by Facebook and hosted by Northern Illinois University, 2021.
  - Advisor, Tarunya Shakti - a Nonprofit Organization doing humanitarian works, 2020 - 2024.
  - Team Leader, RUET_Abnormalz - a Competitive Programming team that participated in 8 National and several International Programming Contests. 2013 - 2017.
  - Secretary (Communication), RUET IT Society (RITS), 2017.
  - General Member (Communication), RUET Programmers' Association (RPA) & ACM Lab, 2014 - 2015.

Certifications

- **IBM AI Engineering Professional Certificate** for successfully completing the following 6 courses:
  - Machine Learning with Python.
  - Scalable Machine Learning on Big Data using Apache Spark.
  - Introduction to Deep Learning & Neural Networks with Keras.
  - Deep Neural Networks with PyTorch.
  - Building Deep Learning Models with TensorFlow.
  - AI Capstone Project with Deep Learning.
- **Machine Learning** - Certificate of completion of an 11-week course with Stanford Online.

Awards and Achievements

- Top 25 Employees: among 1400 employees at Robi Axiata Ltd, Bangladesh for AI Upskilling Program, 2020.
- Top 100 Learners: among 1400 employees at Robi Axiata Ltd, Bangladesh for times spent in online learning, 2020.
- Divisional Champion, National Collegiate Programming Contest (NCPC), Bangladesh, 2015.
- Honorable Mention, ACM ICPC Dhaka regional online preliminary, 2013, 2014, 2015, 2016.
- Honorable Mention, IEEE RoboTour & Programming Contest, RUET, Bangladesh, 2014.
- Honorable Mention, Divisional Math Olympiad (BdMO), Dinajpur, Bangladesh, 2010.
- Best Speaker, Hakimpur Upazila Debate Competition, Bangladesh, 2009, 2010.
- Winner, Hakimpur Upazila Inter School Essay Writing Competition, Bangladesh, 2004, 2010.
- Education Board Scholarships, Bangladesh, 2004, 2007, 2010, 2012.