A S M Shahadat Hossain
1315 W Lincoln Hwy, Apt. 200, DeKalb, IL 60115, USA
asmhossain20@gmail.com (779) 902-0043 www.linkedin.com/in/asmshahadat

SUMMARY

A Data Science enthusiast with 3+ years of industry experience in **Data Science**, **Machine Learning**, **Natural Language Processing**, **Data Warehouse**, **Business Intelligence**, **Customer Insights Platforms**, **Operational Support Systems**, and **DevOps Engineering** at renowned telecommunication companies having **49** to **83 million customers**.

EDUCATION

- M.S. (ongoing) in Computer Science, Northern Illinois University, DeKalb, IL, USA *Expected May 2024*
- B.Sc. in Computer Science & Engineering, Rajshahi University of Engineering & Technology, Bangladesh *Feb 2018*

RESEARCH INTERESTS

Computational Reproducibility, Data Science, Machine Learning, Natural Language Processing, Software Engineering

PUBLICATION [Google Scholar]

- 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.
  [Paper in IEEE XPlore] [PDF] (Number of Citations: 38)

RESEARCH PROJECTS

- **Reproducibility of Visualizations in Jupyter Notebooks: Algorithms vs. Programmers** *[Details]* ***Python***
  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 images.

- **Reproducibility of Jupyter Notebooks** *[Details]* ***Machine Learning, Natural Language Processing, Python***
  Features were extracted from Jupyter Notebooks and **4** machine learning models such as Logistic Regression (**LR**), Extreme Gradient Boosting (**XGBoost**), Bidirectional Encoder Representations from Transformers (**BERT**), and Random Forest (**RF**) were built to automatically assess reproducibility with an accuracy of **62.2%**.

- **'Reproducibility' in CS conferences** *[Details]* ***Machine Learning, Natural Language Processing, Python***
  Texts from **55** Computer Science conference websites mentioning reproducibility and **250** research papers from **55** such conferences as well as **25** conferences not mentioning reproducibility were used to understand how computer science conferences assess reproducibility of the submitted works and if there are any shared characteristics among the papers accepted in reproducibility-aware conferences.

- **Clinical Trials Categorization** *[Details]* ***Machine Learning, Natural Language Processing, Python***
  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).

- **Comparative Study of Clustering Algorithms** *[Details] [Related Publication]* ***Machine Learning, Python***
  A Comparative study of **2** clustering algorithms (**K-means** and **DBSCAN**) based on their performances on **3** datasets (**IRIS** Dataset, **Wholesale Customer** Dataset, and **WEKA** Unbalanced Dataset).

WORK EXPERIENCE

- **Graduate Teaching Assistant,** Northern Illinois University, DeKalb, IL, USA *Aug 2020 - Present*
  - Holding office hours, grading weekly assignments, quizzes, final exams, and proctoring exams for courses:
    CSCI 240: Computer Programming in C++; *Fall 20, Spring 21*
    CSCI 330: UNIX and Network Programming; *Fall 21, Fall 22, Spring 23, Fall 23, Spring 24*
    CSCI 463: Computer Architecture and Systems Organization; *Spring 22*

- **Graduate Staff Assistant (Data Science & BI),** Northern Illinois University, DeKalb, IL, USA *Jun 2022 - Sep 2022*
  - Performed **Data Migration** from flat files to **Snowflake** (**Cloud Computing**-based **Data Warehouse**)
  - **Detected** and **Corrected errors** in **13 Python scripts** caused due to **System Migration** (Windows to Linux)
  - Generated **Data Visualization** using **Microsoft Power BI** for Graduate Career and Professional Development
  - Ran processes on a **Hybrid** CPU/GPU **Computing Cluster** having **60 nodes** capable of **30 teraFLOPS**
  - 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**

- **Specialist (Operational Support Systems),** Robi Axiata Limited, Dhaka, Bangladesh *May 2019 - Feb 2021*
  - Developed at least **10 scripts** using **Python**, **PL/SQL**, **Linux Shell**, **Windows PowerShell**, and **PHP** to automate manual jobs and saved at least **4 man-hours** per day which is approximately **$8K per year**
  - Performed **System Administration** of **Network Management Systems** developed by **Huawei**, **Ericsson**, and **Nokia**
  - Provided **Server Administration** (**Linux**: Red Hat Linux, SUSE Linux, and **Windows**: 2012 R2, 2016) assistance
  - Extended support to the automation Project Managers who used **Agile** and **Scrum** methods
  - Analyzed network performance data of approximately **50 million subscribers** by using tools and scripts

- **Jr. Engineer (Business Intelligence - Data Warehouse),** Grameenphone IT Ltd, Dhaka, Bangladesh *Jul 2018 - May 2019*
  - Performed **Developments** and **Operations** of **ETL** (Extract, Transform, Load) using **SQL** in **Oracle** servers
  - Investigated Data trend Mismatch from revenue up to **$40K per day**
  - Generated monthly High-Value customer database from **83 million subscribers** using **PL/SQL**

- **Project Engineer, Transmission Modernization,** Robi Axiata Limited, Dhaka, Bangladesh *Apr 2018 - Jun 2018*
  - Supported in **Project Management** (Agile) and **Vendor Management**
  - Initiated and monitored at least **10 Network Change Requests** per day for **3G to 4G** network upgradation

SKILLS

- **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 Management**: Enterprise Data Warehousing, Business Intelligence Data Integration, ETL (Extract, Transform, Load) in Oracle, MySQL, and Microsoft SQL Server
- **Data Visualization**: Tableau, Microsoft Excel, Microsoft Power BI, 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: 2012 R2, 2016
- **Interpersonal Skills**: leadership, teamwork, conflict management, patience, and communication

CERTIFICATIONS, AWARDS & ACHIEVEMENTS

- **IBM AI Engineering Professional Certificate** *[Credentials]*
- **Machine Learning** - Certificate of completion of an **11-week course** with **Stanford** Online *[Credentials]*
- **Top 25 Employees:** Among more than **1400 employees** at Robi Axiata Ltd, Bangladesh for **AI Upskilling** Program
- **Top 100 Learners**: Among more than **1400 employees** at Robi Axiata Ltd, Bangladesh for times in online learning
- **Divisional Champion**, National Collegiate **Programming Contest** (NCPC) 2015, Bangladesh
- Honorable Mentions in **ACM ICPC** Dhaka regional online preliminary 2013, 2014, 2015, 2016

SERVICES

- **Volunteer** (Evaluation), **STEM Fest** sponsored by **Meta** and hosted by Northern Illinois University *2021, 2022, 2023*
- **Problem Setter** & **Judge**, Programming contests at RUET **Analytical Programming Lab** *Jan 2015 - Jan 2016*
- Student **Program Committee**, 13th International Conference on the Theory of Information Retrieval **SIGIR** (**ICTIR**) *2023*
- **Reviewer** at The 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), The 3rd International Conference on Electrical, Computer and Energy Technologies (ICECET), International Conference on Artificial Intelligence, Control, Data Sciences and Applications (ACDSA) *2023*