Mahbubur Rahman

+88-01775964184 | durlovrahman32@gmail.com | LinkedIn | GitHub | Mirpur, Dhaka, Bangladesh

Research Interests

Privacy Preserving Data Mining (PPDM)

Data Privacy

Cyber Security

Information Security

Machine Learning

Deep Learning

Artificial Intelligence

Highlights

- Excellent quantitative data research & qualitative analysis skills with Python, MATLAB, SQL and other languages and tools.
- Rich experience in data preparation, statistical technique and data visualization tools.
- Proficiency in programming languages such as: C, C++, Java etc.
- Strong interpersonal, problem-solving, and organizational skills developed in my undergrad life.
- Comprehensive understanding of version control systems (Git workflow and GitHub).
- Solid knowledge from four years' study and research experience in computer science.
- Proficiency in English.

Education

Rajshahi University of Engineering & Technology — Rajshahi, Bangladesh
B.Sc. in Computer Science and Engineering — *Jan 2017 – Oct 2022*
CGPA: 3.35/4.00 *(Last 4 semester: 3.58/4.00)*

Research Experience / Publications

Efficient perturbation techniques for preserving privacy of multivariate sensitive data
Journal: Array — Publisher: Elsevier — Impact factor: 4.5 — Q1 — *Oct 6, 2023*

- Cloud data is increasing significantly recently because of the advancement of technology which can contain individuals' sensitive information, such as medical diagnostics reports. While deriving knowledge from such sensitive data, different third party can get their hands on this sensitive information. Therefore, privacy preservation of such sensitive data has become a vital issue. Data perturbation is one of the most often used data mining approaches for safeguarding privacy. A significant challenge in data perturbation is balancing the privacy and utility of data. Securing an individual's privacy often entails the forfeiture of the data utility, and the contrary is true. Though there exist several approaches to deal with the trade-off between privacy and utility, researchers are always looking for new approaches. In order to address this critical issue, this paper proposes two data perturbation approaches namely NOS2R and NOS2R2. The proposed perturbation techniques are experimented with over ten benchmark UCI data set for analyzing privacy protection, information entropy, attack resistance, data utility, and classification error. The proposed approaches are compared with two existing approaches 3DRT and NRoReM. The thorough experimental analysis exhibits that the best-performing approach NOS2R2 offers 15.48% higher entropy and 15.53% more resistance against ICA attack compared to the best existing approach NRoReM. Furthermore, in terms of utility, the accuracy, f1-score, precision and recall of NOS2R2 perturbed data are 42.32%, 31.22%, 30.77% and 16.15% more close to the original data respectively than the NRoReM perturbed data.
- doi:10.1016/j.array.2023.100324

Work Experience

RTC Hubs Ltd. Dhaka, Bangladesh

Software Engineer

- Developed some comprehensive websites which address customer needs.
- Designed and implemented the front-end interface using React, ensuring a user-friendly and responsive experience.
- Collaborated with the team to integrate and display detailed information.
- Successfully delivered fully functional and visually appealing website that enhanced the user experience.

Projects

Intrusion Detection System (IDS) using machine learning | *Google Colab*

* **Objective**: The goal is to develop a system capable of identifying potential security threats in network communication.
* **Dataset:** The NSL-KDD dataset, a widely used benchmark dataset for intrusion detection
* **Algorithms**: Three distinct machine learning algorithms, namely Support Vector Machine (SVM), Decision Tree, and Naive Bayes, are utilized to explore diverse approaches for intrusion detection.
* **Workflow**: Preprocess data, including one-hot encoding and standardization — Split data into training and testing sets — Evaluate models using metrics like accuracy, precision, recall, and F1 score.
* **Visualizations**: Generated confusion matrices for model performance overview. Accuracy were shown in histograms for three algorithms. Training & Testing time was also illustrated in bar charts.

Cat vs Dog Image Classification Using Transfer Learning | *Google Colab*

* Developed a state-of-the-art animal classification model by leveraging Transfer Learning techniques.
* Used the Kaggle Dogs vs Cats dataset as the foundation for training.
* Employed MobileNet V2, a powerful pre-trained deep learning model, to achieve high accuracy in distinguishing between cats and dogs.
* This project showcases my expertise in machine learning, deep learning, and model deployment

Car Price Predictor Application | *Google Colab* | *GitHub*

* Developed a Car Price Predictor application utilizing machine learning techniques.
* Employed Linear Regression to create a predictive model for estimating car prices based on relevant features.
* Designed and implemented a user-friendly web interface for users to input car details and receive price predictions.
* Leveraged Python, scikit-learn, Flask, and web technologies to build the application.

Heart Disease Prediction using Machine Learning | *Google Colab*

* Developed a predictive model to detect and classify heart disease using machine learning techniques.
* Employed logistic regression, a powerful classification algorithm, to build the predictive model.
* Used Kaggle Heart Disease dataset.

Football Data Analysis & Visualization with Python | *Google Colab*

* Analyzed and visualized multifaceted elements of the English Premier League 2021-22 Season using Python.
* Utilized the EPL22 dataset for this work.
* Delved into various aspects, including top 10 goalscorers, assist providers, player nationalities, total team assists and goals, calculating the average team squad statistics as well as many other aspects of the season.
* The project involved extensive data analysis and visualization techniques, showcasing comprehensive insights into the season's performance and player statistics.

Certifications

* **Supervised Learning with scikit-learn** | Datacamp
* **Unsupervised Learning in Python** | Datacamp
* **Linear Classifiers in Python** | Datacamp
* **Machine Learning with Tree-Based Models in Python** | Datacamp
* **Neural Networks and Deep Learning** | Coursera