SHUTONU MITRA

Falls Church, VA |(929) 682-7538 |Email: mshutonu@vt.edu| LinkedIn: in/shutonu-mitra

Education

Virginia Tech | Virginia, USA   *August 2023 – May 2025*
Master of Science in Computer Science; GPA: 3.83/4.00   Courses: Neural Network, Big Data, Cloud Computing, ML

Military Institute of Science & Technology | Bangladesh   *February 2018 – May 2022*
Bachelor of Science in Computer Science; CGPA: 3.65/4.00

Skills

Languages: Java, Python, C, C++, JavaScript, PHP, SQL, PL/SQL, HTML 5, CSS 3, Bash
Analytic Skill: Data visualization (Tableau, Excel, Python), Machine Learning, Neural Networks (Tensorflow, PyTorch), Natural Language Processing, IBM cloud, AWS cloud services (EC2, EMR, Lamda), PySpark, Google Cloud Platform.
Technologies: Android, Django, MySQL, Firebase, Oracle, Latex, Figma, Open GL, MongoDB, React, Node JS, Rest API.

Experience

Curanostics Inc.   *May 2024– Aug 2024*
Machine Learning Engineer   Suwanee, Georgia, USA

- Collaborating with the product development team to integrate AI features into the platform by developing REST APIs and interactive visualizations of health care data in Next JS deployed in Vercel/Render, enhancing user engagement by 25%.
- Designing and implementing scalable back-end systems for health information technology software, improving performance by 40% and ensuring robust infrastructure through rigorous testing and debugging.
- Leveraging Google Cloud Platform for medical report summarization and question answering, developing and optimizing ML models for healthcare data analysis using Retrieval-Augmented Generation (RAG), SciBERT, BioGPT, Med-PaLM etc.

Selise Digital Platform   *February 2021– March 2023*
Data Scientist   Dhaka, Bangladesh

- Implemented and optimized SQL and NoSQL databases, achieving a 35% performance enhancement through query tuning.
- Automated ETL processes from distributed systems, leveraging Azure Databricks for data storage and Kubernetes for container orchestration. Collaborated on troubleshooting efforts, reducing data pipeline defects by 40%.
- Utilized Python, R and Tableau to analyze, process and visualize large datasets, applying quantitative techniques and optimizing data workflows, making data efficiently accessible for Business Intelligence team.
- Worked with MLOps teams to deploy models in production environments, ensuring continuous monitoring and iterative improvements based on real-time performance metrics, resulting in improving client satisfaction by 19%.

Virginia Tech   *June 2024–Present*
Graduate Research Assistant   Virginia, USA

- Extracted high-quality data from social media for cyber fraud cascade detection, defined social cyber vulnerabilities metrics with faculty, and improved social scam classification accuracy by 25%.
- Developed multimodal data analytics and algorithms, using Graph Neural Network and Reinforcement Learning, applying advanced techniques in NLP and deep learning to analyze unstructured text data, leading to a 20% reduction in false positives.
- Contributed to the creation of a geospatial dashboard that maps SCVI across different regions, identifying high-risk areas.

Projects

Malicious URL Detection | Tensorflow, Pytorch, NLP, Neural Networks   *Git Repository*

- Architected a real-time malicious URL detection system to block phishing, malware, and defacement URLs that lead to malicious websites causing cyber threats by the proposed Convolutional neural network model with an F1 score of 97.67%.

REST API with LAMBDA Functions| JAVA, MySQL, AWS Lambda Function   *Git Repository*

- Devised Developed and managed a suite of AWS Lambda functions to serve as a scalable backend for REST APIs, facilitating seamless data interaction between a MySQL database and a web application frontend.

Predictis | Machine Learning, IoT, Android, Cloud   *Git Repository*

- Created and validated an android application that predicts its user's future possibility of having Cardio-vascular diseases using an ML model built on real-time data from biomedical sensors, improving utility and user satisfaction by 25%.

Twitter Hashtag Counter| PySpark, AWS EMR, AWS S3, Map-Reduce   *Git Repository*

- Devised a custom hashtag analysis pipeline by Map-Reduce algorithm utilizing Apache Spark deployed on Amazon EMR, seamlessly processing Twitter data hosted on AWS S3 to identify and store the top 20 hashtags.

Publications

[1] "A Framework to Detect and Prevent Cyberbullying from Social Media by Exploring Machine Learning Algorithms." International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2) (2021)
[2] "Predictis: an IoT and machine learning-based system to predict the risk level of cardiovascular diseases." BMC Health Serv Res 23, 171 (2023).