Musa J. Taib

403 805 7753 | musajabbartaib@gmail.com | linkedin.com/in/musajtaib/ | github.com/MusaJTaib

Summary

Recent Master's graduate specialized in **Machine Learning and Data Science**, bringing **over three years of research experience** in the field. Currently working as a Machine Learning and Data Researcher, I am actively seeking opportunities to further my career in the fields of Data Science and Machine Learning.

Experience

University of Calgary **Sept 2023 - Present**

Machine Learning and Data Reseacher *Calgary, Canada*

- Developed secure, privacy-focused ML models in **PyTorch** using federated learning with differential privacy for on-site applications, including homelessness trend analysis. This involved advanced time series analysis for data pattern identification in collaboration with local agencies.
- **Authored multiple scholarly papers** and **conference abstracts** and **presented research** at multiple forums to **academic** and **general audiences**.

University of Calgary **Sept 2021 - August 2023**

Teaching and Machine Learning Research Assistant *Calgary, Canada*

- As an RA: Spearheaded a large-scale data analysis initiative by processing and analyzing 200 GB of administrative health data for different databases for both structured and unstructured data, **recommending resource allocation** for over 100,000 clients.
- Executed complex data cleaning and engineering of time series health data into structured temporal matrices for **enhanced feature extraction**, **utilized Seaborn**, **Matplotlib** and **Statsmodel** in Python for intricate EDA and data visualization and used MySQL for robust database management.
- Devised an innovative algorithm for multivariate time series data using dynamic time warping for feature ranking based on effective temporal resolution, reducing matrix sizes by 70% and preserving 95% of model performance.
- Enhanced processing throughput and model training scalability by integrating Compute Canada's ARC HPC capabilities using shell scripting.
- Conducted comprehensive cross-validation on heterogeneous ML models (implemented using **TensorFlow**), achieving a 12% performance boost, thereby affirming the method's robustness and generalizability across varied architectures.
- As a TA: Led labs covering topics from machine learning to network protocols, mentored student project teams towards capstone readiness, administered technical exams, and led undergrad teams with a focus on engineering project management.

Robotics and Machine Intelligence (ROMI) Lab **June 2020 - June 2021**

Data Science Research Intern *Islamabad, Pakistan*

- Mined data and made **NLP models** using PyTorch, integrating GANs to synthesize authentic-feeling network traffic, successfully mimicking patterns from high-traffic platforms like YouTube, Facebook, and Dailymotion.
- Streamlined data analysis and operational efficiency by leveraging **shell scripting in Linux**, coupled with strategic application of clustering algorithms for enhanced anomaly detection and network security insights.

Technical Skills

Languages: **Python**, Java, **C++**, **C**, **Linux shell scripting**, SQL, MATLAB, Latex.
Libraries & Frameworks: Pandas, NumPy, PyTorch, Tensorflow, Keras, SciPy, Matplotlib, Seaborn, Imbalanced Learn, SciKit Learn, Spacy, NLTK, Statsmodels, OpenCV.
Technologies: Linux, Git, MySQL, HPC, AWS, Dockers, SageMaker, Excel, Power BI, ETL.
ML Expertise: **Deep Learning (CNNs, RNNs)**, **Federated Learning**, **Regression**, **Classification**, **Clustering**, **Time series analysis**, **Data Transformation** and **Optimization.**

Education

University of Calgary **Sep. 2021 – Nov. 2023**

Masters of Science in Electrical Engineering *(CGPA: **4.0**/4.0)* *Calgary, Alberta*

National University of Sciences and Technology **Sep. 2017 – Jun. 2021**

Bachelors of Engineering in Electrical Engineering (CGPA: 3.65/4.0)(Minor in Business Studies) *Calgary, Alberta*

Projects

Mortality Prediction using SAPS-2 Features **2022 - Present**

- Orchestrated an ETL pipeline on Amazon S3, managing the vast MIMIC-III database (over 40,000 patients), enhancing data retrieval and storage scalability while ensuring integrity. Employed unsupervised learning for astute feature discernment.
- Engineered and refined predictive models with **GRUs**, **SVMs** and **Random forests**, realizing an AUC-ROC of 85%, demonstrating potent precision in foreseeing patient prognoses within the extensive dataset.

CGANs for Data Augmentation **2022**

- Implemented a **CNN-based Conditional GAN** for data augmentation on MNIST, FashionMNIST and CIFAR-10, generating synthetic images to counteract class imbalance.
- Optimized model performance with HPC-driven parallel job arrays, enhancing F1-Score by 10%.