Usama Baloch

+92 3113074329 | usama.balochhh@gmail.com

LinkedIn | Medium | Kaggle | Github

EDUCATION

Bachelor of Computer Science, National University of Computer and Emerging Science (NUCES) 2019- 2023, *Karachi, PK*

SKILLS

Programming Languages: Python, C, C++, X86 Assembly, and SQL.

Deep Learning & Data Science: PyTorch, TensorFlow, JAX, Keras, Numpy, Scikit-Learn, Pandas, TensorRT, Matplotlib, Seaborn, XGBoost, LightGBM, Transformers, CUDA, TFLITE, OpenCV, Scipy, Cupy, Pytest, Plotly, Jupyter, Data Science Pipeline, Tabular, Time series, Audio, Computer Vision, LLMs, KNN, CNN Architectures, SVM, Linear and Logistic Regression, Decision Tree, Boosting, Bagging, Stacking, PCA, and Statistics.

Version Control Tools and: AWS Cloud (SageMaker, S3, Boto3, Lambda, API Gateway), Git, CI/CD, VSCode, PyCharm, Linux..

Work EXPERIENCE

MLE-I Unify.ai Dec 2022 – Dec 2024, *London, UK (Remote)*

- Achieving **2x-3x speed** in Inference time against torch_tensorrt by revamping **Ivy's TensorRT Backend** using Python API, integrating efficient **CUDA Kernels** and **Python converters**, plus writing testing functions in **Pytest** for the operations in Ivy Framework.
- Implemented optimized Ivy Operations daily using **software engineering** techniques. **Built, trained, and evaluated** AI Models from scratch using Ivy Framework, also maintained AI Models in **production**. Moreover, Collaborated with Project Heads to develop **quarterly roadmaps** and discuss them with **CEO and stakeholders.**
- Write **Documentation** for Ivy Framework using my **writing skills** and Reviewed **GitHub PRs**, resolved bugs, and upheld clean code standards for 5-7 contributors weekly.

Data Scientist Upwork and Kaggle May 2021 – Sept 2022, *Karachi, PK*

- Worked on different data problems like tabular data, time-series, anomaly detection, recommendation system, CV, and NLP.
- Designed complete data pipelines, including **storage, exploratory analysis, feature engineering, feature selection, model selection**, and picking the **right metric** according to the dataset.
- Deployed and Maintained Models in **AWS Cloud** using **AWS SageMaker.** Created **Lambda** functions and **API Gateways** for **Model Endpoints** for **client** applications**.**

PROJECTS

Ivy Framework, Compiler, and Transpiler - GitHub

- Achieved inference of **2x-3x** on Hugging Face Models using Ivy's **TensorRT Compiler** against torch_tensorrt, **Optimized** the Deep learning models using **Sparsification techniques** (specifically post-quantization and quantization aware training).
- Developed Efficient functions and tests using **Software Engineering Techniques** for the Ivy Framework and **documented** the functionalities. Moreover, I Tested **Ivy Transpiler** by converting models and fixing bugs in the Backend to make it stable.

Bike Renting (Time Series) - GitHub

- Achieved **0.289 RMSE** using XGBoost Regressor by designing a **data pipeline** (data cleaning, data visualizing, and feature engineering), including **log transformation** for the out-of-range counts feature. Utilized **AWS SageMaker** to model building, model training, and deployment of the model, saving artifacts in **S3,** and implemented model-specific logic with **AWS Lambda** and **API Gateway** for client access, plus **monitoring** of model and **retraining** it throughout the process.

Optimal Player Substitution Strategy - GitHub

- Developed a Hypothesis to calculate the **Stamina** of football players, achieving **0.93 precision and 0.84 recall** by detecting players using **fine-tuned Yolov5**, tracking and assigning unique numbers to players using **ByteTrack**, and implementing the Stamina Algorithm to those players on **45 mint videos,** also communicated with the supervisor about the progress of the project**.**

Shopee - Price Matching - GitHub

- Developed **eca_nfnet_l0** Deep learning algorithm with **Ranger Optimizer + Mish** Activation to generate Image embeddings to find **Product Similarities** using **Cosine distance**, Used **ArcFace Loss** to ensure similar class embeddings are close and dissimilar ones are far apart, this will help Retailers to find similar products with different prices.
- Achieved **0.71 F1 Score** by combining Image and Text **Embedding Predictions** generated by **eca_nfnet_10 + KNN and TfIdf.**

X-RAY Body Part Classifier - GitHub

- Achieving an **84% accuracy** on the validation set by **normalizing pixel** intensity values for consistent training. Designed and implemented a **CNN Classifier** using TensorFlow to classify different **body parts** from X-ray images in Dicom format.

Bird Sound Classifier - GitHub

- Achieving **0.616 AUC Score** with 5 K-Fold using **EfficientNet-B0** baseline with PyTorch organized with **PyTorch-Lightning.** Accelerated audio pre-processing to mel-spectrograms using **CuPy**, applied **Transformations**(GuassianNoise, PinkNoise, NoiseInjection, etc), and **Augmentations** (Flipping, XYMasking).

Customer Credit - GitHub

- Achieved **80% accuracy rate** by addressing **multi-class classification** on customer bank account data to predict credit categories (Poor, Good, Standard). Engineered a robust **Data pipeline** (EDA, FE, FS), leveraging **XGBoost with cross-validation.**

House Pricing Advanced Regression - GitHub

- Achieved **87% accuracy with XGBoost** Machine Learning algorithm on validation data through **EDA**, including handling NaNs, outliers, and plotting relationships. Used **feature engineering:** statistical groupby, label encoding, and column splitting.

Restaurant Management System

- Created server-side design, and implemented **Advanced SQL queries** to retrieve data from the **client side** and send data from the database to the client side., used **JOIN statements** for faster results. **Created tables** like users, orders, complaints, admin, customer items, and cart items for better representation of the data.

Cervical Spine Classifier - GitHub

- Achieved a **LogLoss metric of 0.54** by employing Transfer Learning with an **Efficient-net-b07 model** with FCNN Attached at last, Utilizing **normalizing** to align the whole dataset on one learning rate and **tiling for image partitioning** to detect cervical spine fractures on CT Scan images and to prevent Ram crashing.

Lung-Cancer Classifier - GitHub

- Achieving a **0.056 CrossEntropy Loss** on the validation set. Group the interesting voxels into **small lumps**, and find the **center point of the nodule** which gives us (index, row, column) locations then combine with **voxel CT voxel data** and feed into the **Custom CNN classifier** written in PyTorch to confirm whether it is a nodule or not.
- Feeded the **classified nodule** into another **CNN Classifier** written in PyTorch to determine whether the nodule is **benign or malignant.** In addition, added **Softmax** at last to know the per-tumor status.

Data Compression - GitHub

- Implemented a research algorithm based on **Huffman Trees** in C++, using problem-solving skills to implement **Octanary and Hexanary Trees** efficiently for data compression, achieving **2x speed** compared to the Binary Tree implementation.

Gaming Zone Project - GitHub

- Utilized **Object-Oriented Programming techniques** in C++ (including Classes, Templates, Filing, and Enum) to develop a system that **tracks records** of all available PCs and Consoles in the gaming zone, **monitors game availability**, and provides an **admin class** with the authority to manage user accounts and game configurations.

Plagiarism Checker X86 Assembly Language - GitHub

- I had **three articles** containing random content. I tried a strategy to **compare the triplets** of article one with article two and **count the frequency** of the **same triplets** if the **frequency was greater than 50%** then it was cheated else it was not. The **formula** for finding the **number of triplets: no of words- 2**, every triplet consists of Three words.
- Used Different sizes of **registers to store and send data** back and forth. implemented **functions** to read the files, find no of triplets in the articles, and compare them. Used **stack frame** for storing the data during calculations.

ACHIEVEMENTS &EXTRA-CURRICULAR ACTIVITIES

- **1) Kaggle Master Highest rank 110 with 6 Golds, 8 Silvers, and 3 bronze medals**
- **2) Open-Source Contribution at TensorFlow, Ivy Framework, and NVIDIA-TensorRT.**
- **3) Winner of Data Science Competition held by DevDay 2023 in NUCES Karachi.**
- **4) Winner of the DataFest Competition held by AppForgers in 2022 at NUCES Karachi.**
- **5) Runner-up in the Data Science Competition held by Procom 2023 at NUCES University of Karachi, got best-visualized recognition.**