Abdul Hanan Khan

abdulhanankhan@web.de ❖ (+49) 1627599058 ❖ Dresden, Germany
❖ https://www.linkedin.com/in/abdul-hanan-khan-230a77228/

WORK EXPERIENCE

AGCO **Nov 2023 – Current**
Data Scientist Intern – Customer Analytics *Marktoberdorf, Germany*

- Leveraging AWS S3 and Athena for robust storage management and comprehensive data analysis, enabling efficient query execution and data retrieval for enhanced decision-making.
- Utilizing BERT, a Large Language Model deployed on Azure Databricks, for in-depth customer comment analysis including topic identification and classification. Additionally, employing Google T5 for sentiment analysis to derive business insights.
- Enhancing model development and performance tracking through the use of Azure MLflow, which is utilized to meticulously monitor BERT and T5 model metrics and improvements over time
- Utilizing AWS Glue to develop crawlers for scheduling automated table generation, seamlessly integrating BERT's output into scalable data storage solutions for downstream analysis and visualization.
- Collaborating with the dashboard team, providing processed data for visualization in Tableau dashboards, facilitating actionable business insights through interactive reporting.

GlobalFoundries **May 2022 – May 2023**
Data Scientist Intern - TCAD *Dresden, Germany*

- Conducted ETL processes on raw transistor data from FEM simulations, ensuring high-quality, structured datasets for analysis and model training.
- Automated hyperparameter optimization for deep neural networks using a blend of Bayesian optimization and random search algorithms, resulting in around 25% performance boost.
- Implemented transfer learning to leverage device physics from one semiconductor device for training neural networks on other devices, leading to reduction in training resources by ~30%
- Generated multi-dimensional parallelized deep neural network implementations through shell scripting, reducing model development time by ~40%.

Bauhaus-Universität **Nov. 2020 – May 2022**
Teacher Assistant (Tutor/HiWi) *Weimar, Germany*

- Developed stochastic simulation techniques in python reducing runtime by almost 30 % compared to previous MATLAB implementation.
- Presented tutorials in Optimization and stochastic simulations.

EDUCATION

Bauhaus-Universität **August, 2024**
M.Sc. Digital Engineering *Weimar, Germany*

- Main subjects: Machine learning, Natural language processing, Image analysis, Computer vision, Software engineering, Algorithms & Data structures
- Thesis: Personalization of LLM's to reduce harmful content generation

PROJECTS

- **Transfer learning in TCAD-enabled machine learning models**
  - Utilized the power of transfer learning to develop an efficient training technique for similar transistor devices using deep NN.
- **Hyperparameter optimization for neural networks**
  - Built a hybrid automated optimization model using Bayesian and random search algorithms for neural network hyperparameters**.**
- **DeeplabV3 background removal: model development and flask deployment on AWS EC2**
  - Deployed a state-of-the-art deep learning model, DeepLabV3, for accurate background removal in images, using Flask on AWS EC2 instance using REST API.
- **Churn Prediction And Comparative Analysis**
  - Conducted exploratory data analysis (EDA), predicted churn through unsupervised KMeans clustering, and validated the results through ANN, Random forest and SVM classifier.
- **Custom deep neural network deployment on AWS sagemaker**
  - Created, trained, and deployed a custom deep neural network on AWS SageMaker using Docker.
- **Low context word prediction with large language models**
  - Fine-tuned language models such as BERT, GPT and n-gram to boost their performance in low-context word prediction scenarios by almost 35%.
- **Web pages classification**
  - Got 1st place in two web page classification competitions using methods like least mean squares, batch gradient descent, and ANN.

- **GITHUB link**
  - https://github.com/HananKhan7/Projects

PUBLICATION

- **TCAD-enabled Machine Learning – An Efficient Framework to Build Highly Accurate and Reliable Models for Semiconductor Technology Development and Fabrication (IEEE, 2023)**
  - Developed highly accurate deep neural networks (Digital twins) using automated hyperparameter optimization and transfer learning for semiconductor technology.

SKILLS

- **Language skills:** English (C2), German (B1)
- **Programming languages***:* Python, Java, Bash , MATLAB
- **Big Data:** SQL (AWS Athena), AWS S3, AWS Glue, Pandas, NumPy, Git
- **Data visualization:** Matplotlib
- **ML libraries:** Tensorflow, Keras, Scikit-learn, PyTorch, OpenCV, NLTK
- **Additional:** Microsoft Office, RESTAPI, FastAPI, Flask