Sundus Aijaz

Data Scientist | ML Engineer | MS Data Science, FU Berlin

sundusaijaz2011@gmail.com | +49 17637665217 | Berlin, Germany | LinkedIn | GitHub | Portfolio

Authorized to work in the EU

PROFESSIONAL SUMMARY

Soon-to-graduate MS Data Science student at FU Berlin (Oct/Nov 2026), targeting entry-level Data Scientist, ML Engineer, or AI Engineer roles. 3 years of hands-on experience building ML systems, statistical models, and data pipelines in Python, PyTorch, TensorFlow, and Scikit-learn, with work in LLMs, time-series forecasting, classification, and applied deep learning. Currently completing an MS thesis at BIFOLD (TU Berlin) on conditional neural network potentials. Fluent in English (C1, working language). German A2 and actively improving.

WORK EXPERIENCE

Working Student, ML Technische Universität Berlin | Berlin, Germany *Dec 2024 – Present*

- Developing a generative model that computationally produces new molecular structures, expanding the training data for ML interatomic potentials.
- Designed and built an ensemble inference system (SpkEnsembleCalculator) in SchNetPack with uncertainty quantification over energies, forces, and stress tensors, supporting active-learning and reliability assessment in atomistic simulations.
- Refactored the core data pipeline, unifying interfaces across 8 molecular and materials datasets (QM9, MD17, MD22, rMD17, MPtrj, ISO17, QM7x, Materials Project) and reducing technical debt across the framework.
- Integrated a large-scale molecular trajectory dataset, giving researchers direct access to a primary training corpus for universal machine-learning interatomic potentials.
- Built pytest test suites across the ensemble, data, and atomistic-offset modules, supporting continuous integration across multiple release cycles.

Software Engineer Lyftrondata | US, Remote *Oct 2021 – Oct 2023*

- Engineered scalable ETL pipelines using Airflow and Kafka, improving processing speed by 30%.
- Collaborated with cross-functional teams in an agile setup to refine data integration strategies.
- Developed custom Python-based connectors for heterogeneous data sources, integrating them into Snowflake and BigQuery pipelines using SQL and PySpark.
- Created interactive Tableau dashboards to visualize key metrics and support business decisions.

AI Engineer Xehen | Karachi, Pakistan *Jul 2020 – May 2021*

- Built web scrapers for 50+ e-commerce websites in Python using Scrapy, Selenium, and BeautifulSoup, storing structured data in PostgreSQL.
- Developed a real-time school attendance system using Hugging Face facial recognition and ensemble ML models for automated identification.
- Built a 96-class dog breed classifier using YOLOv3 and landmark-based features, achieving 97% accuracy.

PROJECTS

Automated Sales Forecasting System *ARIMA · LSTM · AWS QuickSight*

- Designed a time-series forecasting pipeline using ARIMA and LSTM models to predict sales trends; exposed results through AWS QuickSight dashboards to support business decision-making.

Restaurant Footfall Forecasting *ARIMA · XGBoost · Streamlit*

- Built a customer visit forecasting system comparing ARIMA and XGBoost on public datasets, achieving ~15% MAE improvement with XGBoost. Developed a Streamlit dashboard for weekly demand planning.

LLM Engineering *OpenAI API · GPT · LangChain · Prompt Engineering*

- Built a suite of LLM-powered applications including a text summarisation pipeline and a conversational assistant, applying prompt engineering and agentic workflows using the OpenAI API and LangChain.

Face Detection and Tracking Pipeline *MTCNN · OCSORT · Python*

- Built a scalable real-time detection and tracking pipeline using MTCNN and OCSORT; demonstrates applied deep learning and software architecture skills for industrial computer vision contexts.

TECHNICAL SKILLS

Languages: Python, SQL

ML / Statistics: Machine learning, deep learning, computer vision (YOLOv3), Scikit-learn, PyTorch, TensorFlow, XGBoost, time-series forecasting (ARIMA, LSTM), equivariant GNNs (PaiNN, SchNetPack), statistical modelling, feature engineering, uncertainty quantification
LLMs / NLP: NLP, Hugging Face Transformers, OpenAI API, LangChain, prompt engineering, text summarisation
Data Engineering: Pandas, NumPy, Apache Airflow, Kafka, Spark, PySpark, Snowflake, BigQuery, PostgreSQL, MySQL, MongoDB
MLOps / Deployment: MLOps, model deployment, MLflow, Docker, Git, pytest, AWS, Azure
Visualisation: Matplotlib, Seaborn, Tableau, Streamlit, AWS QuickSight

EDUCATION

MS in Data Science Freie Universität Berlin | Berlin, Germany *Oct 2023 – Expected Oct/Nov 2026*
Relevant coursework: Machine Learning, Statistics, Data Management, Data Visualisation, LLMs, Semantic Technologies.
MS Thesis, BIFOLD (TU Berlin): *Transfer Learning for Bridging Machine Learning Force Fields and Diffusion Models*

- Training a single multi-head model jointly on equilibrium-structure and non-equilibrium force datasets to unify ML force fields (MLFFs) and diffusion-based generative models, using a shared representation with separate heads for DFT forces and generative pseudo-forces, evaluated on QM7x against standalone MLFF and GPFF baselines (PyTorch, SchNet/PaiNN).
- Implementing and benchmarking dataset-conditioning strategies (representation and output-shift conditioning) against multi-head readout baselines in PyTorch.

BS in Software Engineering Sir Syed University of Engineering and Technology | Karachi, Pakistan*Jan 2017 – Mar 2021*
Graduated 2nd in cohort, top grade band (1.1, German scale).

LANGUAGES

English (C1, working language) | German (A2, actively improving) | Urdu (native)