

# Aijaz Ali Shah Syed

Data & ML Engineer – Databricks & Fabric Associate

✉ [lakyari.ajjaz@gmail.com](mailto:lakyari.ajjaz@gmail.com) · 📞 +49 157 3302 1370 · 📍 Düsseldorf, Deutschland

Geburtsdatum: 19.10.1999 · Nationalität: Pakistanisch · Visum: Arbeitssuche

[LinkedIn](#) · [GitHub](#) · [Certifications](#)



## PROFIL

Ingenieurshintergrund mit **3+ Jahren Erfahrung in Data Engineering und Machine Learning**. Produktionserfahrung im Aufbau eines **Prognosesystems auf Azure Databricks**, mit vollständiger Verantwortung für den ELT- und ML-Lebenszyklus von der Datenerfassung bis zum Training von Modellen. Erfahrung im Aufbau skalierbarer Datenpipelines und End-to-End-Workflows.

## TECHNISCHE FÄHIGKEITEN

<b>Sprachen</b>	Python, SQL, R, C++
<b>Data Engineering</b>	ETL/ELT Pipelines, Delta Lake, Spark, Airflow, Databricks, Snowflake, Fabric
<b>Cloud &amp; Ops</b>	Azure (ADLS Gen2, ADF, Synapse), Terraform, DABs, CI/CD, Docker, Git
<b>Databricks</b>	Auto Loader, Unity Catalog, DLT/SDP, Databricks SQL, Jobs, Genie
<b>ML &amp; MLOps</b>	PyTorch, TensorFlow, Optuna, scikit-learn, MLflow, Spark Streaming, Model Serving
<b>AI Stack</b>	ChatGPT, Claude (code, cowork), Copilot, n8n, AI Studio
<b>Derzeit erkundet</b>	GCP/ BigQuery, SAP BDC, LLMOps, RAG, Agentic AI

## BERUFSERFAHRUNG

### Werkstudent – Data Engineering & MLOps

10/2023 – 06/2025

Stadtwerke Krefeld AG

- Entwickelte **Azure Databricks** Prognoseplattform für Fernwärme- und Energiesysteme.
- Entwickelte **PySpark & Delta Lake** ELT-Pipelines, arbeitszeit um **30%** gegenüber Benchmark reduziert.
- Orchestrierte **ADF** and **Databricks Asset Bundles** für wiederholbare, produktionsreife ELT-Deployments.
- Implementierte **MLflow** Tracking und Modellversionierung über Multi-Environment **CI/CD** Pipelines.
- Stellte bereit **Spark Structured Streaming** Endpunkte, Time-to-Production von Wochen auf Tage reduziert.
- Verwaltete **Unity Catalog** Metadatenstandards, Datenqualität über die gesamte Plattform sichergestellt.
- Automatisierte **Power BI** Dashboards für Echtzeit-Energieprognosen über Geschäftsbereiche.
- Präsentierte Daten- und ML-Architekturentscheidungen technischen und geschäftlichen Stakeholdern.

### Masterarbeit – Hybride ML-MILP Unit Commitment Optimierung

01/2025 – 06/2025

Stadtwerke Krefeld AG & TU Dortmund

- Entwickelte hybrides **ML-MILP** Framework für Unit Commitment, erreichte **98%** Lösungsoptimalität.
- Reduzierte **Gurobi** Laufzeit um **57,1%** durch Ersetzung von Constraints mit ML Warm-Start-Grenzen.
- Automatisierte **MLflow** Pipeline mit Optuna-Hyperparameter-Tuning und k-facher Kreuzvalidierung.
- Produktionsreife durch Datenvalidierung und Machbarkeitsassertionen bei Modelloutputs sichergestellt.

### Wissenschaftliche Hilfskraft – Forschung

11/2022 – 06/2023

TU Dortmund (ie3 Institut)

- Simulierte 5G-Netzwerkkommunikationsanforderungen für Solarkraftwerk mit **NS-3** and **MATLAB**.
- Analysierte Simulationsergebnisse zur Bewertung von Latenz, Durchsatz und Netzwerkleistungsmetriken.
- Dokumentierte Ergebnisse und unterstützte Forschungspublikationen am ie3-Institut der TU Dortmund.

## Data Engineer Trainee

01/2021 – 06/2022

Torino Technologies, Pakistan

- Verwaltete **SQL** Datenbanken einschließlich Schemadesign, Query-Optimierung und Datenqualitätssicherung.
- Migrierte On-Premises-Datenbank zu **Azure** Cloud für skalierbare und kosteneffiziente Datenspeicherung.
- Baute auf und wartete **ETL/ELT** Pipelines zur Automatisierung von Extraktion, Transformation und Laden.
- Entwickelte automated CI/CD pipelines and data integration scripts using **Python** and **SQL**.
- Lieferte Data-Engineering-Lösungen für internationale und lokale Kunden über mehrere Projekte.

## PROJEKTE

---

### Databricks End-to-End Forecasting Pipeline – [GitHub](#)

- Databricks-basierte Energieprognosepipeline mit **DLT**, **MLflow** und **Unity Catalog**; bereitgestellt über Databricks Asset Bundles (DABs) mit vollautomatisiertem Dev-to-Prod CI/CD.

### Hybrid ML-MILP Optimization Framework – [GitHub](#)

- End-to-End MLOps-Framework, das LSTM & LightGBM mit MILP-Solvern kombiniert — Optuna-Tuning & **MLflow versioning**, k-fold time-series validation.

### Virtual Power Plant Simulation – MSc-Projekt (12 ECTS) – [GitHub](#)

- Simulierte erneuerbare Erzeugung & flexible Lastverteilung mit **VPPLib**; implementierte Betriebsstrategien für kosteneffizientes Verbraucher-Energiemanagement.

### Energy Trading Platform – MSc-Projekt (12 ECTS)

- Agentenbasierte Smart-Factory-Energiegebote mit **Pyomo**; Maschinen bieten für Zeitfenster nach Betriebspriorität & Echtzeit-Energiepreisen.

## ZERTIFIZIERUNGEN

---

- Databricks Data Engineer Associate
- Databricks ML Associate
- Azure Databricks Platform Architect
- Microsoft Fabric DP-700
- Snowflake Essentials (Badge 1 & 2)
- dbt Fundamentals

## BILDUNG

---

**M.Sc. Automation & Robotics** 10/2021 – 09/2025

Technische Universität Dortmund

**Diplom in Technology Management** 04/2023 – 10/2024

Technische Universität Dortmund

**B.E. Mechatronics** 10/2016 – 09/2020

Mehran University of Eng. & Tech.

## SPRACHEN

---

Englisch – C1 (Fließend)

German – B1  
(Aktive Verbesserung)

Urdu – Muttersprache

Sindhi – Muttersprache

## WEITERE INFORMATIONEN

---

**Hobbys:** Sport, Reisen, Kochen

**Führerschein:** Deutscher Klasse B