

PARSA KAMALI

M.Sc. Data Science Student | Cloud Data Engineering | Python, SQL, AWS & Databricks

Bochum, Deutschland | parsa.kamali.data@gmail.com | [linkedin.com/in/parsa-kamali](https://www.linkedin.com/in/parsa-kamali) | github.com/Parsaka
Portfolio & vollständige Projektdokumentation: parsa-kamali.de

PROFIL

M.Sc. Data Science Student an der TU Dortmund University mit praktischer Erfahrung in Datenanalyse, Datenqualitätsbewertung, Python, SQL und akademischen Datenprojekten mit großen Datensätzen und Machine-Learning-Workflows. Fokus auf Data Engineering und Cloud Data Engineering mit praxisorientiertem Lernen in AWS, Databricks, Spark und ETL/ELT-Pipeline-Design. Offen für Working-Student-/Werkstudentenpositionen in Data Engineering, Cloud Data Engineering oder Data Analytics in Deutschland.

TECHNISCHE SKILLS

Programmierung & Abfragen: Python, SQL, Bash-Grundlagen

Datenverarbeitung & Analyse: Pandas, NumPy, Polars, Jupyter Notebook, Data Cleaning, Data Preprocessing, explorative Datenanalyse, Feature Engineering, Datenvisualisierung

Data Science & Machine Learning: Scikit-learn, XGBoost, PyTorch, BERT-basierte Textverarbeitung, Supervised Learning, Klassifikation, Modellevaluation, Precision/Recall/F1, Confusion Matrix

Data Engineering: ETL/ELT-Konzepte, Datenpipelines, Batch Processing, Datenqualitätsbewertung, Datentransformation, Grundlagen der Datenmodellierung, Grundlagen der Workflow-Orchestrierung

Cloud & AWS: Amazon S3, AWS Glue, Amazon Athena, Amazon Redshift, AWS Lambda, AWS Step Functions, IAM, CloudWatch, KMS

Databricks & Big Data: Databricks, Apache Spark, PySpark, Delta Lake, Lakehouse-Architektur, Lakeflow Jobs

Graph Analytics: PyTorch Geometric, Graph Neural Networks, GAT/GATv2, Reply-Graph-Modellierung, Edge Classification

Tools & Entwicklung: Git, GitHub, Docker-Grundlagen, Linux-Grundlagen, VS Code, HTML, CSS, Bootstrap, PHP-Grundlagen

AUSBILDUNG

M.Sc. Data Science | TU Dortmund University | Dortmund, Deutschland | Aktuell

- Relevante Bereiche: Datenanalyse, Machine Learning, Statistik, Data Mining, Simulation-Based Inference, Computational Intelligence und graphbasierte Modellierung.

B.Sc. Software Engineering | Golestan University | Iran

BERUFSERFAHRUNG

Data Analyst | TELUS Digital AI Data Solutions | Essen, Deutschland | Mai 2022 - Heute

- Bearbeitung von Aufgaben in Datenanalyse und Qualitätsbewertung in einem internationalen AI-Data-Solutions-Umfeld.
- Überprüfung, Analyse und Bewertung von Datensätzen zur Erkennung von Mustern, Inkonsistenzen und Datenqualitätsproblemen.
- Arbeit nach strukturierten Workflows und Qualitätsstandards zur Unterstützung zuverlässiger datengetriebener Prozesse.
- Aufbau einer ausgeprägten Genauigkeit, Konsistenz und Zuverlässigkeit bei datenbezogenen Aufgaben.

Software Engineer | Poshtibanan Pardazesh Aseman | Gorgan, Iran | 2017 - Juni 2020

- Backend-Entwicklung für datenbankgestützte Web- und Mobile-Application-Projekte, hauptsächlich mit PHP, SQL und Microsoft SQL Server.
- Entwurf, Pflege und Abfrage relationaler Datenbankstrukturen zur Unterstützung von Geschäftsprozessen, Nutzerdaten und Anwendungsfunktionen.
- Entwicklung von Backend-Logik zur Verbindung PHP-basierter Anwendungskomponenten mit SQL-Server-Datenbanken und strukturierten Datenoperationen.
- Arbeit mit SQL-Abfragen, Datenabruf, Datenaktualisierung und Backend-Anwendungslogik für produktionsorientierte Software-Anwendungsfälle.
- Aufbau einer starken Backend- und Datenbankgrundlage durch praktische Erfahrung in Debugging, Problemlösung und Projektauslieferung.

AUSGEWÄHLTE PROJEKTE

Telegram Message Deletion Analysis Pipeline | Unabhängiges Data- & Graph-Analytics-Projekt | Python, Polars, PyTorch Geometric, XGBoost

- Entwicklung einer Python-basierten Datenverarbeitungs- und Graph-Analytics-Pipeline zur Analyse von Telegram-Nachrichtenlöschungen.
- Bereinigung und Verarbeitung großer Nachrichtendatensätze mit User-IDs, Zeitstempeln, Reply-Beziehungen, Topics und Löschlabels.
- Erstellung strukturierter Features aus zeitlichen Mustern, Nutzerverhalten, Reply Chains, Nachrichtenmetadaten und graphbasierten Interaktionen.
- Training und Evaluation von Modellen zur Löschvorhersage mit XGBoost und Graph-Neural-Network-Ansätzen.
- Erzeugung analysefähiger Outputs, Evaluationsmetriken und visueller Zusammenfassungen für Modellvergleich und Interpretation.

AWS & Databricks Lakehouse Pipeline | Unabhängiges Portfolio-Projekt | AWS, Databricks, Spark, Delta Lake, ETL/ELT

- Konzeption einer End-to-End-Cloud-Data-Engineering-Pipeline auf Basis von AWS- und Databricks-Lakehouse-Konzepten.

- Planung von Bronze-, Silver- und Gold-Datenebenen für Raw Ingestion, Data Cleaning, Transformation und analysefähige Outputs.
- Praxis mit AWS-Services wie Amazon S3, AWS Glue, Amazon Athena, Amazon Redshift, IAM, AWS Step Functions, CloudWatch und KMS.
- Anwendung von Databricks-Konzepten wie Apache Spark, PySpark, Delta Lake, Lakehouse-Architektur, Lakeflow Jobs und Workflow-Orchestrierung.
- Fokus auf skalierbare Datenverarbeitung, Datenqualitätsprüfungen, Partitionierungskonzepte, reproduzierbare Transformationen und jobrelevante Cloud-Data-Engineering-Workflows.

Eye-Tracking Data Analysis Case Study | Akademisches Datenanalyseprojekt | Python, Data Cleaning, EDA, Visualisierung, Statistical Reporting

- Analyse von Eye-Tracking-Fixationsdaten zur Untersuchung teilnehmerbezogener Muster visueller Aufmerksamkeit.
- Arbeit mit fixationsbasierten Messwerten, Teilnehmerinformationen und experimentellen Datenstrukturen.
- Durchführung von Data Cleaning, explorativer Datenanalyse und strukturierter Interpretation verhaltensbezogener Muster.
- Erstellung analytischer Outputs, visueller Zusammenfassungen und schriftlicher Berichte auf Basis experimenteller Ergebnisse.

Personal Portfolio Website | Unabhängiges Web-Portfolio-Projekt | PHP, HTML, CSS, Bootstrap, GitHub

- Entwicklung einer persönlichen Portfolio-Website zur Darstellung technischer Projekte, CV, Skills, Kontaktinformationen und beruflichem Profil.
- Aufbau strukturierter Bereiche für Profil, Erfahrung, Projekte, Skills und Kontakt.
- Nutzung von PHP, Bootstrap, HTML, CSS, GitHub und grundlegenden lokalen Entwicklungstools für Implementierung und Präsentation.

Portfolio-Hinweis: Die oben genannten Projekte sind ausgewählte Zusammenfassungen. Vollständige Projektdokumentation, Screenshots, zusätzliche Projekte und unterstützende Materialien sind auf meiner Portfolio-Website verfügbar: parsa-kamali.de

ZERTIFIZIERUNGEN & TRAINING

- AWS Certified Data Engineer - Associate (DEA-C01) - in Vorbereitung
- Databricks Certified Data Engineer Associate - Prüfungsvorbereitung
- Databricks Academy Training mit Fokus auf Lakehouse-Architektur, Spark, Delta Lake und Data-Engineering-Workflows

SPRACHEN & WEITERES

Sprachen: Persisch: Muttersprache | Englisch: C1 | Deutsch: B1, aktuell Verbesserung Richtung B2

Führerschein: Klasse B