

Rakesh Nagaragatta Jayanna

rakeshnagaragattajayanna@gmail.com | +491777752620 | Mannheim | LinkedIn | GitHub | Portfolio

PROFILE

Data professional with 3+ years of experience in data engineering, machine learning, analytics, and applied AI. Skilled in building scalable data workflows and practical solutions using Python, SQL, Spark, and cloud platforms.

AUSBILDUNG

Masters in Applied Data Science and Analytics, 10.2023 – 09.2025 | Heidelberg, Germany
SRH University Heidelberg

Bachelors in Computer Science and Engineering, 06.2016 – 09.2020 | Hassan, India
Malnad College of Engineering

BERUFSERFAHRUNG

AI Development Intern, *iiterate Technologies GmbH* 05.2025 – 04.2026 | Koblenz, Germany

- Built a GenAI-based internal search assistant by cleaning and structuring 1,000+ enterprise documents for reliable retrieval.
- Implemented a RAG pipeline using OpenAI APIs, Hugging Face embeddings, and vector search (FAISS, Pinecone) to improve answer relevance and reduce retrieval errors.
- Delivered PoCs and demos using LangChain and Streamlit, adding caching, basic monitoring, and iterative prompt and retrieval tuning to improve user satisfaction by 15%.

Analyst, *KPMG Global Services Private Limited* 11.2020 – 08.2023 | Bengaluru, India

- Developed and maintained Azure Data Factory pipelines that ingested data from the Policy Administration System (Majesco) and external APIs into Azure Data Lake Storage improving ingestion throughput by 30%.
- Performed large-scale data preprocessing in Azure Databricks using PySpark and Spark SQL including joins, data cleaning, deduplication and schema standardization for fraud analytics use cases.
- Optimized end-to-end pipeline and reporting performance through partitioning, caching, better file sizing and Power BI model optimization, reducing processing time by 35% and database load by 30%.
- Implemented data quality and reconciliation checks (row counts, null checks, key uniqueness) and automated failure notifications to improve pipeline reliability and reduce downstream reporting issues.

PROJEKTE

Azure End-to-End Data Engineering Project *ℙ*, *Tech stack: Azure Data Factory, Azure Data Lake Gen2, Azure Databricks, Azure Synapse Analytics, Power BI, PySpark, SQL, Git, Azure DevOps, Docker*

- Built an end-to-end Azure data pipeline using Data Factory, Data Lake Gen2, Databricks, Synapse and Power BI, implementing a Medallion Architecture to ingest, transform and model AdventureWorks data for near real-time reporting and business-ready insights.
- Supported CI/CD and deployment workflows using Git, Azure DevOps and Docker to enable version control, pipeline testing, and more consistent delivery across development stages.

Azure AI-Powered Document Extraction Pipeline *ℙ*,
Tech stack: Python, Azure Durable Functions, GPT-4 Vision, Azure Container Apps, Azure Monitor, Docker, CI/CD

- Built a scalable document extraction pipeline for PDF processing using Azure Durable Functions with batch-based orchestration.
- Used GPT-4 Vision with structured prompting to extract key fields from invoices and convert them into consistent outputs.
- Deployed the service with Docker and Azure Container Apps, with monitoring, logging, retries, and CI/CD to improve reliability and extraction efficiency by 25%.

SKILLS

- Programming Languages: Python, SQL, Java
- AI & Data Solutions: LLMs, RAG, Prompt Engineering, Vector Databases
- Data Engineering & Cloud: ETL/ELT, Apache Spark, Databricks, Data Pipelines, AWS, Azure
- DevOps & Deployment: CI/CD, Docker, Kubernetes, Monitoring
- Software Engineering & Integration: API-Based Architectures, Agile Teamwork, Stakeholder Communication

SPRACHEN

- English || German

CERTIFICATIONS

- Fundamentals of Deep Learning *Issued by NVIDIA* || Microsoft Certified: Azure Data Engineer Associate