ATHARVA HATEKAR
Data Scientist | AI Engineer

📞 +49 15758905132   ✉ atharva_hatekar@yahoo.in   📍 Cottbus, Germany
🔗 atharvahatekar   🐙 atharvahatekar   🌐 Portfolio

Data Scientist & AI Engineer with 3+ years of experience working on industrial, time-series, and unstructured data. Strong expertise in unsupervised anomaly detection, predictive maintenance, and production-grade Generative AI systems. Proven ability to build end-to-end machine learning solutions—from data preprocessing and feature engineering to deployment using modern MLOps practices.

Work Experience

🚩 **Metso,** Frankfurt, Germany

Master Thesis   11/2025 – Present
Topic: Unsupervised Anomaly Detection on High-Dimensional Multivariate Industrial Time Series Data.
Designed and benchmarked unsupervised anomaly detection models (PCA, Autoencoders, GDN, SimAD) on high-dimensional industrial sensor data.
Developed an Adaptive PCA-based detection approach to handle concept drift and dynamic operating conditions in real-time process data.
Performed extensive data preprocessing, feature engineering, and collaborated with process engineers and domain experts to qualitatively validate detected anomalies and reduce false positives.

Data Scientist   04/2025 – 10/2025
Developed and productionized time-series outlier detection and forecasting models to support predictive maintenance use cases.
Implemented CI/CD pipelines for model deployment and monitoring, and integrated MLflow for experiment tracking, model versioning, and performance comparison across multiple outlier detection approaches.
Built real-time monitoring dashboards using Plotly and gRPC for live data streaming and outlier visualization.

Data Science Intern   10/2024 – 03/2025
Conducted exploratory analysis on multivariate industrial time series data to identify sensor dependencies and abnormal patterns.
Evaluated multiple forecasting and outlier detection techniques to assess feasibility for predictive maintenance.
Built interactive visualization tools using Plotly to help engineers explore complex sensor relationships.
Defined baseline data cleaning rules, KPIs, and validation strategies in collaboration with R&D and domain experts.

🚩 **Dreamwarez,** Pune, India

Data Analyst   08/2021 – 02/2023
Designed and maintained PowerBI dashboards with DAX-based KPIs to support executive decision-making and operational insights.
Performed data cleaning, transformation, and exploratory analysis using SQL, Pandas, and Python.
Applied statistical analysis and machine learning techniques to uncover trends and patterns in structured and semi-structured datasets.
Delivered data-driven insights to stakeholders, translating analytical results into actionable business recommendations.

Software Developer Intern   02/2021 – 07/2021
Developed user-centric web applications with a focus on data validation, backend stability, and responsive design.
Collaborated in Agile/Scrum teams to implement features, debug issues, and optimize application performance.
Built robust input validation pipelines to prevent invalid data ingestion and reduce backend processing errors.

Education

🔖 **Master of Science in Artificial Intelligence**   04/2023 – Present
Brandenburgische Technische Universität Cottbus, Germany.
Key Coursework: Data Warehousing, Mathematical Statistics, Data Mining, Information Retrieval, Image Processing, Computer Vision.

Education (continued)

Bachelor of Technology in Computer Science and Engineering 06/2017 – 06/2021
Lovely Professional University, Jalandhar, India.
Key Coursework: Data Structures & Algorithms, DBMS, Software Engineering, Engineering Mathematics, Web Programming.

Technical Skills

- **Programming:** Python, SQL, C++

- **Machine Learning:** PyTorch, TensorFlow, Scikit-learn, XGBoost, LightGBM

- **Generative AI:** LangChain, LangGraph, RAG, Hugging Face, LlamaIndex

- **MLOps & Backend:** MLflow, Docker, Kubernetes, Azure ML Studio, GitHub Actions (CI/CD), FastAPI

- **Data & Databases:** Pandas, PostgreSQL, MySQL, FAISS, ChromaDB

- **Visualization & BI:** PowerBI, Plotly, Matplotlib, Streamlit

Projects

Document Intelligence Hub 06/2025
Developed a RAG-based document intelligence system enabling semantic search and Q&A over complex technical documents.
Implemented FAISS for efficient similarity search on document embeddings and FastAPI for real-time ingestion and inference.
Designed modular pipelines for document parsing, embedding generation, and response synthesis.

E-Commerce Product Intelligence Engine 03/2025
Built a product intelligence and recommendation system by transforming raw e-commerce data into vector embeddings.
Enabled similarity-based product comparison and discovery using LLM-assisted retrieval techniques.
Developed a modular web scraper and Streamlit application to support end-to-end data ingestion and analysis.

QuizCrafter AI 01/2025
Developed an AI-driven quiz generation system producing context-aware MCQs and True/False questions using LLMs.
Analyzed user performance trends using visual analytics to evaluate learning progress and question difficulty.
Implemented strict output validation using Pydantic to ensure structured and reliable model responses.

Equity News Researcher Tool 07/2024
Built a news retrieval system to support equity research by mapping user queries to relevant financial news articles.
Applied embedding-based similarity search using FAISS to improve relevance and information discovery.
Designed a Streamlit interface to enable interactive exploration of retrieved insights.

Customer Churn Prediction 01/2024
Conducted exploratory data analysis using plotly and data preprocessing using Pandas on customer data.
Trained and evaluated multiple machine learning classification models to predict customer churn, achieving 91% F1 score.
Analyzed feature importance to identify key drivers of customer attrition and support retention strategies.

Research Publication

"Passenger security using facial emotion recognition system"
International Journal of Advance Research in Engineering, Science & Management (IJARESM), Sep. 2022

Languages

- **English** Professional proficiency (C1)

- **German** Elementary Proficiency (A2)