# Aniket Phutane

## Data Scientist | Machine Learning Engineer

Berlin, Germany
(+49) 15906433766
aniketdphutane@gmail.com
PORTFOLIO
LINKEDIN
GITHUB


## SKILLS

| | |
|---|---|
| **Programming languages** | Python |
| **Technical Skills** | Predictive Modeling \| Statistical Analysis \| Time-Series Forecasting |
| **DevOps/Cloud Computing** | Docker \| MLOps (Git, MLflow, FastAPI) \| Amazon Web Services Ecosystem |
| **Databases/Frameworks** | PostgreSQL \| Pytorch \| Apache Spark \| RESTful API \| Databricks |

## EXPERIENCE

### Helmholtz Zentrum Berlin – Data Scientist — Berlin, Germany — Mar 2024 – Ongoing

- Engineered expertise in a niche domain (XAS spectra) by analyzing research papers, building domain knowledge, and identifying trends in machine learning applications.
- Overcame limited labeled data by leveraging open-source datasets, implementing advanced data augmentation (Gaussian noise, interpolation), and deploying beta-VAE for synthetic data generation.
- Developed a CNN1D model with contrastive learning to predict properties from XAS spectra, achieving a **test-set prediction error of <1%**.
- Effectively communicated results and methodologies to both technical and non-technical stakeholders.
- Developed and deployed a user-friendly dashboard using Docker and Streamlit.

### BASF – Machine Learning (Working Student) — Ludwigshafen, Germany — Mar 2022 – Oct 2023

- Investigated Graph Neural Networks (GNNs) as a more effective alternative to traditional 1D NLP methods for molecular data, highlighting their potential to capture spatial relationships.
- Implemented GNNs for corrosion prediction in metals and biodegradation prediction for small molecules and polymers.
- Achieved **annual savings of €120k** by developing a more accurate and efficient solution for biodegradation prediction, surpassing commercial tools.
- Enhanced model interpretability (via SHAP) and robustness by incorporating confidence scores and uncertainty estimation.
- Collaborated with cross-functional teams to deploy the solution via Docker and FastAPI, achieving widespread adoption (6 teams) and securing a patent.

### E.ON - Research Assistant — Aachen, Germany — May 2021 – Oct 2023

**Predictive Maintenance in Industrial Systems**

- Mastered the analysis of complex multivariate time series data for advanced predictive analytics.
- Created a prototype for predicting the Remaining Useful Life (RUL) of aircraft turbo engines using the NASA CMAPSS dataset, achieving an **RMSE of less than 7%.**
- Collaborated with a multidisciplinary team to develop a production-ready product utilizing Flask and Docker, showcasing full-scale deployment capabilities.
- Adapted the predictive model using transfer learning for deployment on a distinct dataset, demonstrating flexibility across varied data environments.

**Power Grid Optimization using Multi-agent Deep Deterministic Policy Gradient (MADDPG)**

- Developed a MADDPG architecture for multi-agent reinforcement learning (MARL), featuring two specialized actors and a global critic to optimize distinct grid sections. Each actor learned to manage its assigned area, while the global critic evaluated the collective performance.
- Implemented optimization strategies, including epsilon-decay for balanced exploration-exploitation, memory replay for efficient experience storage, and target networks with soft updates to improve learning stability and convergence.
- Integrated rules-based systems to enhance decision-making and complement the MARL model, improving overall system performance.

- Demonstrated the effectiveness of the MARL system in identifying critical grid vulnerabilities and enhancing decision-making through comprehensive A/B testing, showcasing improved performance metrics.
- Presented and defended the research findings in a comprehensive master's thesis to academic faculty and industry stakeholders, effectively communicating the methodology, results, and key insights.

**Vodafone - Data Scientist** **Pune, India** **July 2020 – Dec 2020**

- Engineered and managed SAP-HANA systems, designing custom SQL procedures and functions to enhance data retrieval efficiency and streamline processing workflows.
- Developed a tailored data model optimized for Celonis integration, creating a process mining dashboard that identified the bottom 20% cumulative spending patterns across 45 countries for a global client.

**Teras Energies - Data Scientist** **Mumbai, India** **July 2018 – June 2020**

- Extracted and transformed MongoDB wind turbine datasets into Pandas dataframes to enable efficient exploration and analysis of over 5 million data points across multiple tables.
- Developed a SARIMA-based multi-variate time series model and implemented rule-based and isolation forest anomaly detection techniques, addressing class imbalances with SMOTE, to identify critical engine failure points.
- Designed and deployed over 20+ interactive Tableau dashboards, clearly conveying insights from predictive modeling and anomaly detection to onsite teams and clients.

## EDUCATION

**RWTH Aachen University** **Aachen, Germany**
M.Sc., Data Science - Transcript **2021 – 2023**

## PATENT, PROJECTS & TECHNICAL ARTICLES

### Patent

- **DEEP NEURAL NETWORKS FOR BIODEGRADABILITY**, Inventor, BASF, 2024

### Projects

DICOM Harmonizer – **LLM Healthcare Hackathon Winner at UKSH** - View on Github

- Developed an end-to-end DICOM Image Analysis pipeline integrating a Flask backend and Streamlit frontend, enabling secure upload, processing, and visualization of medical images.
- Implemented automated metadata extraction and LOINC code harmonization for DICOM files, leveraging machine learning via LangChain's Ollama integration to analyze image attributes such as modality, body part, and protocol.
- Optimized API design with Flask CORS and structured endpoints.

Novel Materials Discovery (NOMAD) Q&A System – **LLM Hackathon Winner at CSMB** – View on Github

- Designed a tailored question-answering system for NOMAD, leveraging Hugging Face LMs and embedding techniques.
- Created a vector store for document indexing and integrated a Streamlit-based user interface for seamless interaction.

### Technical Articles

- Authored technical articles on RAG Evaluation - From Theory to Implementation, Evaluate Multimodal Models – A Comprehensive Guide, Expectation–Maximization Algorithm Demystified and Understanding the Attention Block.

## ACHIEVEMENTS

- AWS Certified Cloud Practitioner: Verification link (Validation number: 510ET7Y1ZBB41694).
- Gold Medalist: International Olympiad of Mathematics, 2011 - Top 0.1% performance with 50,000+ participants.