Aniket Phutane

Machine Learning Engineer

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

SKILLS

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

EXPERIENCE

Helmholtz Zentrum Berlin – Machine Learning Specialist — Berlin, Germany — Mar 2024 – Ongoing

- Developed domain expertise in XAS spectra by analyzing research papers, leveraging open-source datasets, and applying advanced ML techniques for trend identification.
- Built a CNN1D model with contrastive learning for property prediction, achieving <1% test-set error, and used beta-VAE for synthetic data generation to overcome labeling constraints.
- Communicated insights to diverse stakeholders and deployed interactive dashboards via Docker and Streamlit for seamless accessibility.

Tech Stack: Python, PyTorch, Scikit-learn, MLflow, Docker, Streamlit, Contrastive Learning, Beta-VAE, CNN1D.

BASF – ML Engineer — Ludwigshafen, Germany — Mar 2022 – Oct 2023

- Developed Graph Neural Networks (GNNs) based models for corrosion and biodegradation prediction, improving accuracy and surpassing commercial tools, leading to €120k annual savings.
- Enhanced model interpretability with SHAP and uncertainty estimation, increasing stakeholder trust and adoption across six teams.
- Deployed scalable solutions via Docker and FastAPI, contributing to a patented innovation in molecular property prediction.

Tech Stack: Python, PyTorch Geometric, SHAP, FastAPI, Docker, GNNs, Databricks, Spark, Uncertainty Estimation.

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

Predictive Maintenance in Industrial Systems

- Developed a Transformer-based model for degradation prediction of aircraft turbo engines, achieving an RMSE below 7%.
- Optimized model adaptability with transfer learning, ensuring deployment across diverse datasets and operational conditions.
- Built and deployed a production-ready solution using Flask and Docker, demonstrating end-to-end implementation expertise.

Tech Stack: Python, PyTorch, Transformer Models, Spark, Databricks, Flask, Docker, Time Series Analysis, Transfer Learning.

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

- Developed a MADDPG-based multi-agent reinforcement learning (MARL) system with specialized actors and a global critic, optimizing decision-making in power grids.
- Enhanced model stability and convergence using epsilon-decay, experience replay, and soft target updates.
- Integrated rule-based systems to complement MARL, improving grid vulnerability detection and decision-making.

Tech Stack: Python, PyTorch, Reinforcement Learning, MADDPG, Grid2Op, NumPy, Scikit-learn.

Vodafone – Data Scientist — Pune, India — July 2020 – Dec 2020

- Designed and optimized SQL procedures in SAP HANA to improve data retrieval efficiency, integrating Celonis for process mining and identifying cost-saving opportunities across 45 countries.

Tech Stack: SAP HANA, SQL, Celonis, Process Mining, Data Modeling.

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

- Developed a SARIMA-based multivariate time series model with anomaly detection (Isolation Forest, rule-based methods) to identify critical wind turbine failure points, addressing class imbalances with SMOTE.

Tech Stack: Python, Pandas, MongoDB, SARIMA, Isolation Forest, SMOTE, Tableau.

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

- Built an end-to-end DICOM Image Analysis pipeline leveraging LLMs for automated metadata extraction, modality classification, and protocol standardization, improving processing efficiency.
- Integrated LangChain's Ollama with Flask and Streamlit to enable interactive image analysis, enhancing accessibility for medical professionals.

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.