

# Anup Kumar

MACHINE LEARNING RESEARCHER · PHD

Badweg 7, 79189 Bad Krozingen, Germany

☎ (+49) 1522-6169-642 | ✉ anup.kumar.9187@gmail.com | 🗣 anuprulez | 🌐 anup-kumar-phd-9ab935159/

## Summary

I am a machine learning (ML) researcher in the computer science department at the University of Freiburg, Germany. I finished PhD in computer science and applied machine learning in September 2024. I am looking for a new opportunity in the field of ML and data science, where I am interested in building end-to-end scalable applications based on machine learning, deep learning, and cloud technologies.

## Work Experience

### University of Freiburg

Freiburg, Germany

POST-DOCTORAL RESEARCHER

September 2024 - Present

- Node prediction in undirected graph using graph neural networks and explaining predictions using explainable AI
- Building scientific Q&A system by fine-tuning (LLM) (Llama2 and Mixtral) and using retrieval augmented generation (RAG) to better responses
- Developing model inference tools for open-source machine learning models from model zoos
- Over 400 successful pull requests on GitHub
- Machine learning instructor training more than 60 budding ML scientists

### University of Freiburg

Freiburg, Germany

RESEARCHER

September 2018 - September 2024

- Next token prediction using NLP-based deep learning approaches (recommendation systems). The model inference is available to more than 100,000 researchers assisting them in creating data processing pipelines
- Open-source JupyterLab colaboryatory for training ML models and sending training jobs to remote computing clusters. Researchers can create, train, and evaluate any ML model and manage ML projects
- Open-source and web-based ML tools for GPU-enabled cloud computation. These tools enable researchers to train and evaluate ML models in a few clicks using a web-based platform on a large computing cluster

### Galaxy project (European Galaxy server)

Freiburg, Germany

RESEARCH ASSISTANT (HiWi)

May 2016 - August 2018

- Developed fuzzy search for Galaxy tools
- Migrated Mako files to VueJS

### Pramati Technologies

Hyderabad, India

SENIOR SOFTWARE DEVELOPER

December 2012 - March 2015

- Led a team of 3 front-end developers to create user interface integrations for Gallagher Insurance
- Customised several web controls built in JQuery and Angular UI

### MAQ Software (Microsoft and Envision Communications)

Hyderabad, India

SOFTWARE DEVELOPER

July 2011 - November 2012

- Developed stored procedures in SQL using Microsoft's sales data for business intelligence projects
- Led a group of 4 developers for creating and delivering web application for Envision Communications
- Microsoft certifications: Querying Microsoft SQL Server 2012, SQL Server 2008, Business Intelligence Development and Maintenance and TS: Microsoft SQL Server 2008, Database Development

### ProcessMAP Technologies (now Ideagen EHS)

Hyderabad, India

SOFTWARE ANALYST

July 2010 - July 2011

- Developed and migrated web pages written in classic ASP to ASP.Net for environmental, health and safety (EHS) commercial product

## Skills

### Machine learning/deep learning

Pytorch, TensorFlow, Scikit-learn: Ensemble methods, Transformer, RNN, GNN, XAI, LLMs, LoRA, RAG

### Programming

Python, Matlab, XML, HTML, LATEX, Bash, Ubuntu (Linux), Vim, VSCode

### Data science

Jupyter notebooks, ONNX, NumPy, Pandas, Matplotlib, Seaborn

### MLOPs

Docker, AWS, GitHub Actions, Amazon Sagemaker

### Open-source ML

Web-based ML tools, training and evaluating ML models on Linux-based GPU platforms, Conda recipes

### ML Trainer

Leading a deep learning course at the University of Freiburg

### Languages

English (Expert), German (Intermediate), Hindi (Native)

## Education

### University of Freiburg

PHD IN COMPUTER SCIENCE AND APPLIED MACHINE LEARNING.

- Grade: Magna Cum Laude (Sehr Gut/ Very Good)
- Thesis: <https://doi.org/10.6094/UNIFR/257331>

Freiburg, Germany

September, 2018 - September, 2024

### University of Freiburg

MASTER'S IN COMPUTER SCIENCE

- Grade: 1.7 (Good)
- Thesis: [https://github.com/anupruez/latex\\_master\\_thesis/blob/master/master\\_thesis.pdf](https://github.com/anupruez/latex_master_thesis/blob/master/master_thesis.pdf)

Freiburg, Germany

April, 2015 - August 2018

### National Institute of Technology

BACHELOR IN INFORMATION TECHNOLOGY

- Grade: 7.6 / 10
- Secured 3643 (within 0.01%) rank in the country-wide college entrance examination (2006) out of over 500,000 candidates

Kurukshetra, India

July, 2006 - June, 2010

## Publications

- **Anup Kumar**, G. Cuccuru, B. Gruening, and R. Backofen. *An accessible infrastructure for artificial intelligence using a Docker-based JupyterLab in Galaxy*. GigaScience. 2023. DOI: 10.1093/gigascience/giad028.
- **Anup Kumar**, B. Gruening, and R. Backofen. *Transformer-based tool recommendation system in galaxy*. BMC Bioinformatics. 2023. DOI: 10.1186/s12859-023-05573-w.
- **Anup Kumar**, H. Rasche, B. Grüning, and R. Backofen, *Tool recommender system in Galaxy using deep learning*. GigaScience. 2021. DOI: 10.1093/gigascience/giaa152.
- Q. Gu, **Anup Kumar**, S. Bray, et al., *Galaxy-ML: An accessible, reproducible, and scalable machine learning toolkit for biomedicine*. PLOS Computational Biology. 2021. DOI: 10.1371/journal.pcbi.1009014.
- The Galaxy Community. *The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update*. Nucleic Acids Research. 2022. DOI: 10.1093/nar/gkac247.
- The Galaxy Community. *The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update*. Nucleic Acids Research. 2024. DOI: 10.1093/nar/gkae410

## Machine learning tutorials

### Galaxy Training Network.

<https://training.galaxyproject.org/training-material/topics/statistics/>

ML TUTORIALS

- A Docker-based interactive Jupyterlab powered by GPU for artificial intelligence in Galaxy
- Age prediction using machine learning
- Classification in Machine Learning
- Fine tune large protein model (ProtTrans) using HuggingFace

Online (Open-source)

September, 2018 - Present

## Certifications

### Machine learning in Production

COURSERA

- <https://coursera.org/verify/PQQSRXD6YWCN>

Online at Coursera

December, 2024

### Microsoft SQL Server and Business Intelligence Certifications

MICROSOFT CERTIFICATIONS

- Querying Microsoft SQL Server 2012 (Certification number: E075-3711)
- Microsoft SQL Server 2008, Business Intelligence Development and Maintenance (Certification number: E019-4937)
- TS: Microsoft SQL Server 2008, Database Development (Certification number: D531-3300)

Microsoft Certified Professional

2011-2012