Rabin BK

Data Scientist | Trustworthy ML for clinical data | LLMs, time series, evaluation
M.Sc. Data Science, FAU Erlangen-Nürnberg · Erlangen, Germany

✉ rabinbk588@gmail.com
+49 176 8079 8613
rabinbk.com.np
GitHub: Crazz-Zaac | LinkedIn

Education

Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) — *Erlangen, Germany*
M.Sc. Data Science (final grade pending) — *2023 – 2026*

- **Thesis:** *Structured Extraction of Retrospective MIMIC-IV Studies* (submitted 23 Nov 2026) | Thesis
- **Supervisors:** Prof. Dr. David B. Blumenthal, Dr. Christel Sirocchi

Tribhuvan University — *Kathmandu, Nepal*
B.Sc. Computer Science and Information Technology, grade: 73.98/100 (First Division) — *2016 – 2020*

Research experience

Structured Extraction of Retrospective MIMIC-IV Studies — *FAU, Erlangen*
Master's thesis, Biomedical Network Science Lab; supervised by Prof. Dr. David B. Blumenthal and Dr. Christel Sirocchi | Code
2026

- Designed a 38-question extraction schema covering cohort, outcome, time windows, features, preprocessing, modelling and evaluation of clinical studies.
- Built a fully local retrieval-augmented LLM pipeline: section-aware chunking of text and tables, hybrid dense + BM25 retrieval in Qdrant, and quantised Llama 3.2 and MedLlama 2 models served via Ollama on a 6 GB GPU.
- Created reference answers for 12 studies and a custom manual-review tool; attributed errors to retrieval or generation and validated an automated LLM-as-judge evaluation (RAGAS) against the manual review, with bootstrap confidence intervals.

Investigating Informative Missingness in Time-Series Laboratory Data — *FAU, Erlangen*
Bioinformatics lab project (10 ECTS), supervised by Dr. Christel Sirocchi | Code — *2025*

- Studied whether missing laboratory measurements carry clinical information by predicting missingness not at random (MNAR) in MIMIC-IV laboratory time series for aplasia and neutropenic-fever cohorts.
- Built cohorts and missingness features, compared XGBoost, CatBoost, Random Forest, Gradient Boosting and logistic regression, and ran the pipeline reproducibly with Docker and Apptainer on HPC.

Nepali Number Plate Recognition using Support Vector Machine (SVM) — *Tribhuvan University, Nepal*
B.Sc. final-year project (team of four), College of Applied Business; supervised by Mr. Kamal Acharya — *2021*

- Collected vehicle videos at parking sites, extracted frames and annotated number plates and characters with LabelImg (5,846 labelled character images).
- Built a pipeline with YOLOv4-tiny for plate detection and character segmentation, followed by an SVM classifier for 12 character classes (digits 0–9 and the zonal codes BA and PA).
- Developed a Streamlit web interface for image, video and live-stream input.

Work experience

Primetals Technologies Germany GmbH — *Erlangen, Germany*
Working Student – Data Science / Machine Learning — *Dec 2024 – Present*

- Work on multiple data science projects related to computer vision, alongside time-series analytics for industrial sensor data.
- Benchmarked anomaly-detection workflows on 35+ open time-series datasets with differing sampling rates and characteristics.
- Designed a two-stage motif search: SAX-based candidate filtering, then exact subsequence matching (STUMPY) on high-resolution sensor data.
- Built Python workflows for data preprocessing, labelling, model training and evaluation for an in-house real-time anomaly-detection system analysing industrial sensor data.

Forschungszentrum Jülich – Helmholtz Institute Erlangen-Nürnberg (HI ERN) — *Nuremberg, Germany*
Working Student — *May 2024 – Jan 2025*

- Developed an inventory and privilege-management system for laboratory experiments, with controlled user access and automated database backups.
- Connected JupyterHub to the experiment database (MySQL) so researchers could submit, replicate and track experiments.

Lumbini Academy — *Chuchhepati, Kathmandu, Nepal*
Computer Lecturer, +2 level (grades 11–12) — *Dec 2021 – Feb 2023*

- Delivered lectures on programming, algorithms and database fundamentals; guided practical lab sessions.
- Supervised student projects; taught 150+ students across 3 annual shifts (25+ students per batch).

Wiseyak Inc. *Kathmandu, Nepal*
Data Science Intern (B.Sc. internship) *Jul 2021 – Oct 2021*

- Wrote graph queries (GraphQL/DQL) on HL7 FHIR data for the clinical dashboard and automated discharge-summary reports of WiseMD, an AI-supported electronic medical records platform; implemented follow-up classification logic in Clojure.

Lyceum Model School *Nepal*
Computer Science Teacher, grades 7–10 *May 2015 – Jul 2022*

- Taught computer science to secondary-school students in grades 7–10.

Software projects

Yugantar – Internal Fund Management System
Full-stack web application *2025 – Present*

- Policy-driven cooperative management platform (FastAPI, PostgreSQL/Alembic, React, Docker) with role-based deposits, loans and administrative workflows.

Gele Trekking *Live website*
MERN-stack web application *2025 – Present*

- Mentored a student on database schema and architecture; led further development and Docker-based deployment.

Skills

Programming	Python (advanced), SQL, Bash, LaTeX
Machine learning	PyTorch, TensorFlow, scikit-learn, HuggingFace, OpenCV
LLMs & retrieval	Ollama, Qdrant, BM25 and dense hybrid retrieval, RAGAS, Langdock
Data & time series	MIMIC-IV, HL7 FHIR, sensor time series, missing-data analysis, anomaly detection, STUMPY, SAX
ML systems	Docker, Apptainer, Slurm (HPC), Git, GitHub Actions, GitLab CI/CD, FastAPI, Flask, Streamlit, LabelStudio

Languages

English (C1), German (B1), Nepali (native)