Pankaj Chejara

Portfolio | pankajchejara23@gmail.com | +372 58102666

EDUCATION

PH.D., LEARNING ANALYTICS
Tallinn University
Aug 2024 | Tallinn, Estonia

MTECH, COMPUTER ENG.
Malviya National Institute of Technology
June 2012 | Jaipur, India

MCA, COMPUTER APP.
Rajasthan Technical University
June 2010 | Kota, India

BSC, MATHEMATICS
University of Rajasthan
May 2007 | Jaipur, India

SKILLS

MACHINE LEARNING & AI
Deep Learning, Multimodal ML, Bayesian Inference, Predictive Modeling, A/B Testing

PROGRAMMING & DEVELOPMENT
Python, PyTorch, R, Scikit-learn, Pandas, FastAPI, NumPy, PyTorch, Matplotlib, SciPy, Pgmpy, C++ (working knowledge)

SYSTEMS & PIPELINES
Real-time ML Services, Snakemake, MLflow, Airflow, Scalable Data Pipelines

DATA & CLOUD
SQL, Large-scale log/stream data, GCP (familiar)

RESEARCH IMPACT
Research Publications, Conference Presentations

COLLABORATION & BUSINESS IMPACT
Cross-functional team leadership, Experiment Design, Data-Driven Decision Making

LINKS

pankajchejara23
pankaj-chejara-5ab77855
pankajchejara
@pankajchejara

EXPERIENCE

METROSERT AS | Data Scientist
Sep 2024 – Present | Tallinn, Estonia

- Developed and maintained an **end-to-end genomics data pipeline** (Snakemake, Python, Shell), enabling **scalable, reproducible, and automated processing of high-volume datasets**.
- Delivered **predictive ML models (Random Forest)** for colorectal cancer risk prediction from microbiome data, **achieving 84% diagnostic accuracy**.
- Authored technical documentation and presented results to cross-functional teams, supporting **data-driven decision-making in healthcare research**.
- Contributed **to data governance policy design**, ensuring compliance, security, and consistent practices across workflows.
- **Leading** a cross-functional team of data scientists and data analyst, driving project delivery and meeting project's objectives.

TALLINN UNIVERSITY | Researcher / Developer (Ph.D. Work)
Sep 2019 – Aug 2024 | Tallinn, Estonia

- Applied **machine learning and deep learning** on **multimodal collaboration data (audio, video, logs)** to predict learning outcomes and real-time interaction quality.
- Built **real-time collaboration monitoring systems** (FastAPI microservices, dashboards) that processed video, audio, and log data for actionable feedback.
- Designed **scalable data pipelines** for ingesting and preprocessing large event logs, video, and audio datasets (200+ classroom sessions).
- Collaborated with interdisciplinary teams, resulting in **award-winning ML systems** adopted in real-world classrooms.

IIITM | Researcher
Aug 2016 – June 2018 | Gwalior, India

- Conducted **large-scale network analysis** experiments on social media data, advancing community detection and information diffusion prediction.
- Developed **graph-based ML models** for forecasting engagement patterns, combining classical ML and network science approaches.
- Built **video engagement data pipelines** integrating YouTube API with Moodle for learning analytics applications.

SELECTED PROJECTS

Collaboration Quality Prediction System
Best EdTech Demo Award – Learning Analytics & Knowledge Conf., USA, 2023

- Designed and deployed a **real-time web app for collaboration monitoring**, integrating multimodal ML models for audio/video/log-based predictions.

Genome sequencing data processing workflow

- Built **automated data pipelines** for gut microbiome data for downstream analysis