

Stuti Nayak

The Hague, Netherlands | [Portfolio](#) | stuti257@gmail.com | +31 (0) 686 122823 | [linkedin.com/in/stuti-nayak](https://www.linkedin.com/in/stuti-nayak)

Data Scientist (NLP + Speech) with 4+ years building ML solutions in regulated contexts. Strengths in evaluation/experiment design, reproducible ML (MLflow), and production deployment (Docker, Cloud Run), plus data reliability (Kafka, data contracts).

Work Experience

Data Engineer, Aurelius Enterprise, Amsterdam, Netherlands April 2025 – Present

- **Built** an OAuth2-secured event ingestion prototype for **IoT telemetry**, evaluating Kafka as a scalable alternative to MQTT and documenting production readiness criteria (security, throughput, failure modes).
- **Co-designed** integration patterns and **data contracts** (schema/versioning + validation) to reduce breaking changes and improve reliability for downstream analytics consumers.
- **Contributed** to Aurelius Atlas (open-source governance/metadata): lineage + dataset contracts + access-aware analytics to improve dataset discoverability and reuse in compliant environments.

Data Scientist, Knownwell, The Hague, Netherlands February 2024 – July 2024

- **Built a candidate-selection/routing model** using transformer embeddings (BERT) + BERTopic to assign items to the best-matching topic, improving topic-match accuracy **20% to 30%** and speeding routing/insight extraction.
- Developed **RoBERTa** sentiment classifier for business communications; validated on held-out data and delivered insights for decision-making.

Data Scientist, Rate.nl, Rotterdam, Netherlands August 2021 – September 2023

- Delivered an **audio analytics pipeline** using **wav2vec 2.0** to assess call quality and outcomes, increasing model accuracy from **27% to 55%**.
- Designed experiment plan and **metrics**, tracked runs in **MLflow**, performed **statistical comparison/error analysis** to select production model.
- Deployed containerised inference using **Docker** on **Google Cloud Run** (manual deployment), supporting operational usage.

Researcher, Clinical Data Science, Maastricht University, Netherlands September 2019 – August 2020

- Configured infrastructure/pipelines for **Vantage (federated learning)**, enabling cross-hospital analytics **with privacy guarantees**.
- Designed a **semantic data model** to standardise hospital data and reduce preprocessing overhead for collaborative studies.

Research Assistant, Institute of Data Science, Maastricht University, Netherlands July 2017 – February 2019

- Implemented **Biomedical Metadata Quality Assessment pipeline** with deep learning over publication abstracts to improve data **findability and reusability**.
- Built a **multi-label classifier** (CNN & BiLSTM); **CNN achieved 83% accuracy**.

Skills

Languages: Python, R

ML & Stats: scikit-learn, PyTorch/TensorFlow, XGBoost, NumPy, SciPy; evaluation design; calibration; cross-validation

Data & Scale: SQL, Kafka, ELT/ETL, PySpark, Pandas, dimensional modelling, data quality checks

MLOps: MLflow (experiment tracking), Docker, CI/CD (GitHub Actions)

Cloud/DBs: GCP, Azure, BigQuery, PostgreSQL

Certifications: [Confluent Certified Developer for Apache Kafka](#), [Certified Data Management Professional - Associate](#)

Education

Maastricht University Maastricht, Netherlands

M.Sc. in Data Science in Decision Making

Coursework: Model Identification and Data Fitting, Data Mining, Algorithms for Big Data, Mathematical Optimisation, Stochastic Decision Making, Information Retrieval and Text Mining, Planning and Scheduling, Ontology Engineering and Semantic Web.

Dhirubhai Ambani University Gandhinagar, India

B.Tech. in Information and Communication Technology

Coursework: Probability and Statistics, Data Structures, Design and Analysis of Algorithms, Data Mining and Warehousing, Database Management Systems, Web Data Management, Computer Networks, Software Engineering, Systems Software, Logic for Computer Science.

Publications

- **Stuti Nayak**, Amrapali Zaveri, Pedro Hernandez Serrano, and Michel Dumontier. “Experience: Automated Prediction of Experimental Metadata from Scientific Publications.” *In J. Data and Information Quality* 13, 4, Article 21 (December 2021), pp 1–11.
- **Stuti Nayak**, Amrapali Zaveri, and Michel Dumontier. “Quality assessment of biomedical metadata using topic modelling.” *In 2nd Workshop on Semantic Web Solutions for Large-Scale Biomedical Data Analytics (SeW@BMeDA’18)*. 2018.