

Aniket Mishra

DATA SCIENTIST · DATA ENGINEER · MACHINE LEARNING ENGINEER

☎ (+31) 06 87325708 | ✉ anmishraofficial@gmail.com | 🏠 Aniket-Mishra.github.io | 📺 aniket-mishra | 📱 aniket97

Education

Eindhoven University of Technology (TU/e)

MSC IN DATA SCIENCE AND ARTIFICIAL INTELLIGENCE

September 2024 – September 2026

SRM Institute of Science and Technology

B.TECH IN COMPUTER SCIENCE

May 2016 – June 2020

Experience

Avathon (formerly SparkCognition)

DATA SCIENTIST

March 2021 – July 2025

- Designed, developed, deployed, and maintained robust real-time ETL pipelines for over **10GW** of energy farm assets feeding a SaaS platform, utilizing Pandas, Django, PostgreSQL, Serverless (AWS Lambda), ElasticSearch, Airflow, Docker, and microservices architecture.
- Built and enhanced regression and normal behavior models to predict solar/wind turbine downtime **2-4 months** in advance, leveraging domain-specific clustering (DBC), rolling mean features, and actual vs. predicted analysis.
- Optimized AWS Lambda/Container ETL processes for space/time efficiency, reducing costs by **48%**.
- Improved API performance, and achieved **>300%** faster InfluxDB I/O, contributing to over **70%** operational cost reductions.
- Developed custom Streamlit applications for model performance tracking and analysis, live KPI validations, hyperparameter tuning, and developer QoL products.
- Engineered an anomaly detection system for wind and solar assets with **>95%** accuracy in capturing device failures and minimizing false alarms; implemented as a platform feature and a scheduled database flagging job.
- Optimized historical Elastic data pull runtime, from 4+ hours per asset for 6 months of data to 15 minutes.
- Designed and built a centralized Python library for renewable energy code using Poetry and PyTest for packaging/testing, deployed internally via JFrog/PyPi.
- Onboarded, trained, and mentored **multiple** new hires on team processes, codebase, and data science methodologies, supervising their initial project work.

Infosys

SYSTEMS ENGINEER – DATA AND ANALYTICS

October 2020 – March 2021

- Oversaw ETL operations for **2000+** workflows, administering mappings, sessions, and workflows within Informatica PowerCenter.
- Supported client database and service migrations, ensuring **zero data loss** during transitions.
- Achieved **highest scores in 2** data engineering capstone projects during intensive training program.

Codezoned

CORE TEAM DEVELOPER

June 2018 – Present

- Led development team contributions to multiple open-source projects, including 'ScriptsDump' (large public repository for CS/Math scripts).
- Served as Technical Team Lead on the Organizing Committee for HAC-2020 (MLH Member Hackathon).

Freelance

DATA SCIENCE INSTRUCTOR

August 2023 – August 2024

- Delivered corporate training focused on practical Data Analytics and Machine Learning for experienced professionals.
- Successfully mentored participants seeking career transitions, guiding them through skill acquisition and job search strategies, resulting in **multiple hires** into analytics roles.

Publications

JETIR Conference

SRFBGAN: SUPER-RESOLUTION FEEDBACK GAN

May 2020

- Developed a novel GAN generator architecture incorporating feedback loops, implemented using PyTorch for modeling and PIL for image processing.
- Achieved average improvements of **11%** in PSNR and **6%** in SSIM compared to contemporary state-of-the-art super-resolution models.
- Received **Best Research Paper award** at RCICD 2020 conference.
- **Link:** JETIR Paper

Projects

Third Eye – Threat Detection System

HAC'KP-20 HACKATHON PROJECT

Aug 2020

- Built a threat detection prototype using OpenCV (facial-recognition) and SQLite to identify faces against a simulated police database.
- Utilized an Nvidia Jetson Nano for edge processing and designed a real-time alert mechanism.

Heart Failure Severity Detection System

SRM IST (COLLABORATION WITH AIIMS BHUBANESWAR)

Aug 2018

- Developed a system to classify heart failure severity (Levels 1-4) using a Sequential Neural Network (Keras).
- Collaborated with hospital doctors to integrate clinical insights regarding input signals, enhancing model relevance.
- Employed Pandas for data manipulation and Scikit-Learn for preprocessing tasks.
- **Link:** GitHub Repository

Technical Skills

LANGUAGES

Python, SQL, HTML, CSS, C, C++, JavaScript, Rust

DATABASES & CLOUD

PostgreSQL, MySQL, InfluxDB, AWS, GCP, Docker

LIBRARIES & FRAMEWORKS

Pandas, Scikit-Learn, Pytorch, Seaborn, Plotly, Flask, Django, Numpy, Multiprocessing, Streamlit, Dash, Transformers, MLX

TOOLS

Git, Ollama, Jira, Confluence, Swagger, Postman, Excel, PowerPoint, jFrog

Conferences & Hackathons

BEST RESEARCH PAPER AWARD

RCICD Conference - Image Super Resolution using Feedback Loops in GANs (May 2020)

TECHNICAL TEAM LEAD, ORGANIZING COMMITTEE

HAC-2020 MLH Member Hackathon (Apr - Sep 2020)

TOP QUALIFIER

HAC'KP 2020 Hackathon (Aug 2020)

WINNER (1ST POSITION)

Robotics Competition (Arduino) - IIT Madras (Mar 2018)

Certifications & Specializations

NPTEL (CERTIFICATION)

Programming, Data Structures, and Algorithms in Python

UNIVERSITY OF MICHIGAN (SPECIALIZATION)

Statistics with Python Specialization