Aniket Mishra

Data Scientist & Data Engineer

Eindhoven, NL • +31 6 8732 5708 • a.mishra3@student.tue.nl

linkedin.com/in/aniket97 • github.com/Aniket-Mishra • aniket-mishra.github.io

Professional Summary

Data Scientist and Data Engineer with 4.5 years of experience in designing **ETL pipelines, predictive models, and anomaly detection systems** for **10GW+ renewable energy assets**. Proven track record in reducing cloud costs by **55%**, accelerating analytics pipelines by **16×**, and capturing 98% of false and missing alarms. Skilled in Python, AWS, and advanced machine learning, with research interests in **LLMs, AutoML, and optimizations**. Experienced in both **academic research** and **industry-scale AI systems**.

Education

Eindhoven University of Technology (TU/e), Eindhoven, NL   Sep 2024 – Sep 2026 (expected)
MSc, Data Science and Artificial Intelligence

- **Majors:** Data Mining & Machine Learning; Data Engineering & Management
- **Minor:** Process Mining & Visual Analytics
- **Relevant coursework:** Machine Learning Engineering, Deep Learning, Process Mining, Visual Analytics, Survival Analysis, Foundations of Artificial Intelligence, Machine Learning for Industry, Big Data Management, Knowledge Engineering, Research Topics in Data Mining, Philosophy and Ethics of AI, Engineering Data Intensive Systems

SRM Institute of Science and Technology, Chennai, IN   Jun 2016 – Sep 2020
B.Tech, Computer Science *(First Class with Distinction)*

- **Thesis:** *Image Super-Resolution using Feedback Loops in GANs*
- **Best Research Paper Award:** RCICD Conference 2020
- **Relevant coursework:** Python Programming, Artificial Intelligence, Database Management Systems, Pattern Recognition & Expert Systems in Medicine, Neuro-Fuzzy & Genetic Programming, Data Science & Big Data Analytics, Enterprise Resource Planning, Marketing Management, Digital System Design, Object-Oriented Analysis & Design, Algorithm Design & Analysis, Software Engineering Principles

Professional Experience

Avathon (formerly SparkCognition), Bengaluru, IN   *Mar 2021 – Jul 2025*
Data Scientist / Data Engineer

- Scaled real-time ETL pipelines for **10GW+** renewable assets, reducing API latency, onboarding **40+ assets**, and cutting cloud costs by **55%**.
- Built anomaly detection system configurable by asset/rule, **reducing false alarms by 98%** and **capturing 92%** of missed alarms.
- Developed predictive downtime models (2–4 months in advance) with **>95% accuracy**.
- Authored Python ETL framework adopted company-wide; **cut customer onboarding from 3 months to 1 day**.
- Optimised ElasticSearch queries from **4+ hours to ∼15 minutes per asset**.
- Optimized AWS Lambda and containers, lowering compute and storage costs by **48%**; reduced site data model memory usage by **50%**.
- Deployed Streamlit apps for KPI validation, hyperparameter tuning, and monitoring; **used by 20+ stakeholders**.
- Mentored and onboarded **5+ engineers**, improving project delivery timelines.
- Improved MCR processing accuracy to **99% OEM parity**.

Infosys Ltd., Mysore/Hyderabad, IN   *Jan 2020 – Mar 2021*
Systems Engineer, Data & Analytics (incl. Internship)

- Internship top performer; completed **2 capstone projects** (Banking & Insurance).
- Managed **2000+ ETL workflows** in Informatica PowerCenter with **99.9% uptime**.
- Executed database migrations with **zero data loss**.

Data Science Instructor (Freelance), Remote   *Aug 2023 – Sep 2024*

- Delivered corporate training in analytics, ML, and engineering to professionals.
- Mentored learners into analytics roles with portfolio projects and interview prep.

Phoenix Software Solution, Bhubaneswar, IN   *May 2018 – Jun 2018*
Web Developer Intern

- Developed GST billing and tax-invoice web platform using PHP/MySQL.
- Built responsive frontend using Bootstrap, JavaScript, and AJAX.

Research & Publications

- **Best Paper Award** — Research Conference on IoT, Cloud & Data Science (RCICD 2020). Mishra, A. *et al.*, "**SRFBGAN: Image Super-Resolution using Feedback Loops in GANs**," *JETIR*, 2020. Link

Projects

- **Third Eye – Threat Detection (Kerala Police Prototype)**   OpenCV, Jetson Nano
  Real-time facial recognition prototype integrated with police DB; deployed on edge hardware. Repo
- **PDF Workbench**   Python, Streamlit
  Browser-based PDF toolkit (merge, filter, reorder, extract); optimized for real workflows. Repo
- **Reinforcement Learning Robot Navigation**   PyTorch (PPO, DQN)
  Implemented multiple RL algorithms for autonomous robot delivery; PPO outperformed others. Repo
- **Competitive Sudoku AI**   Monte Carlo, MCTS, Heuristics
  Built AI agents for competitive Sudoku. Benchmarked against multiple algorithms including Monte Carlo Tree Search and Markov Chains. Repo
- **Movie Recommendation & Analytics System**   Dash, XGBoost
  Interactive recommender with clustering, PCA, and live dashboards. Repo
- **Historical Shark Attacks Visualization**   Dash/Plotly
  Interactive visualization of shark attack datasets with time and geo-analysis. Repo
- **Heart Failure Severity Detection**   Keras, Healthcare AI
  NN classifier (levels 1–4) built with doctors at AIIMS. Repo
- **dotfiles: Dev Productivity Toolkit**   Rust, Zsh, Bash
  Automations and environment setup for macOS/Linux (CLI tools, shell functions, fast setup scripts). Repo
- **scripts: Automation Utilities**   Python, Rust, OpenCV
  Small tools for everyday workflows: PDF/image ops, inactivity prevention, batch file tasks, and CLI helpers. Repo
- **statistical-model-implementer**   Python, ML
  Prototype library to auto-apply/stat-evaluate models across datasets, compare metrics, and surface the best performer. Repo
- **quickSilver: Graph Path Query Estimator & Engine**   C++17, CMake
  Designed and implemented a cardinality estimator and fast path query evaluation for a C++ in-memory graph database (directed, labeled RDF-style graphs; simple regular path queries). Project

Full archive available on GitHub and Portfolio Website.

Leadership & Volunteering

- **Codezoned (NPO), Core Team & Developer (2018–Present):** Moderated multiple large open-source repositories (incl. *ScriptsDump*, a major CS/Math script archive); managed community servers and maintained the organization's website.
- **Hack Against Covid 2020 (MLH, Devfolio)**: Technical Team Lead: Led infra automation, registration system, and community management for **1000+ participants** from **15+ countries**.
- **ALERT (Alertathon)**: Organizing Volunteer for the event "Run to save a life," Chennai (2018).

Awards & Competitions

- **Best Research Paper Award**, RCICD 2020.
- **Winner (1st)**, IIT Madras Robotics (Arduino), 2018.
- **Top Qualifiers**, HAC-KP 2020 Hackathon.

Skills

- **Programming:** Python, SQL, Rust, JavaScript, C++, Bash
- **ML/AI:** PyTorch, Transformers, AutoML, Anomaly Detection, Time Series Forecasting, Reinforcement Learning, GNNs, Generative AI, Stable Diffusion
- **Data Engineering & MLOps:** Airflow, Docker, Serverless, CI/CD, Terraform
- **Databases & Search:** PostgreSQL, MySQL, InfluxDB, ElasticSearch, DuckDB, Neo4j, ChromaDB
- **Cloud:** AWS, GCP, Heroku, Netlify
- **Visualization:** Plotly, Dash, Streamlit, Panel
- **Tools:** Git, Swagger, Postman, Jira, Confluence, Ollama, ComfyUI, Poetry, UV, Warp

Certifications

- **Statistics with Python Specialization** (Coursera), 2020
- **Programming, Data Structures & Algorithms with Python** (NPTEL), 2019
- **Python Bootcamp: Zero to Hero** (Udemy), 2018

Additional Information

- **Languages:** English (Fluent), Hindi (Fluent), Odia (Native), German (Beginner), Japanese (Beginner)
- **Work Eligibility:** EU student residence permit
- **Interests:** Research in LLMs, AutoML, and data-intensive systems; community hackathons and open-source projects
- **Teaching:** Delivered corporate training on ML & Data Engineering; mentored professionals into analytics roles
- **Hackathons:** Participant and mentor at 5+ hackathons; winner at IIT Madras Robotics, top qualifier HAC-KP
- **Open Source:** Contributor and maintainer on multiple GitHub projects
- **Creative Interests:** Writing poems; using storytelling and pictures to explain concepts

References available on request.