Ritwik Mishra

Bilaspur, India | mritwik11@gmail.com | +91-7693876033
linkedin.com/in/ritwikm4 | github.com/ritwik4m

Education

MSc Data Science (Distinction)   *Sept 2024 – Nov 2025*
University of Surrey, Guildford, UK

- **Modules:** Machine Learning, NLP, Cloud Computing, Data Mining, Business Analytics, Database Systems
- **Dissertation:** *BioDynaMo-Based Pipeline for Neuron Morphology Optimisation* — Developed a Python-based simulation workflow using BioDynaMo to model and optimise neuron structures at scale. (Distinction)

BEng Electronics & Telecommunication   *Aug 2014 – Jun 2018*
Bhilai Institute of Technology, India

- Bachelor's Capstone project: Designed a wireless IoT network for polyhouse monitoring and control.

Technical Skills

- **Data Science & Machine Learning:** Supervised/Unsupervised Learning, NLP, Transformers (LLMs, VLMs, SLMs), Deep Learning (PyTorch, TensorFlow), Scikit-learn, Feature Engineering, Model Evaluation, Prompt Engineering, LLM Evaluation, AI Safety (research-driven), Experimentation Design, A/B Testing (conceptual), Statistical Testing
- **Python Libraries:** Pandas, NumPy, Scikit-learn, Transformers, TensorFlow, PyTorch, Matplotlib, Seaborn, SpaCy, Flask, FastAPI
- **Programming Languages:** Python, SQL, Java, C/C++, JavaScript
- **Data Engineering & Analytics:** ETL, Data Validation, Data Quality Checks, Structured/Unstructured Data, BigQuery, Power BI, Tableau, Anomaly Detection, Forecasting
- **Cloud & DevOps:** Azure (incl. Azure ML Studio), AWS, Google Cloud Platform, Docker, Kubernetes, CI/CD Pipelines
- **Systems & Integration:** REST APIs, JSON, XML, Distributed Systems, Middleware Integration
- **Workflow & Collaboration:** Git, GitHub, VS Code, Jupyter, PyTest, Postman, Agile/Scrum, Stakeholder Communication, Reporting

Experience

Paxcom India (A Paymentus Company)   *Gurgaon, India*
QA Engineer – Data Validation & Analytics   *May 2022 – Aug 2024*

- Developed Python-based **data validation pipelines** and **SQL test automation** for financial datasets, enhancing pipeline reliability and reducing manual QA by 30%.
- Built **Power BI dashboards** for QA metric reporting across financial products, improving stakeholder visibility and audit readiness.
- Contributed to **experimentation efforts** (A/B-style testing) to evaluate the impact of validation rules on downstream analytics accuracy.
- Collaborated with data engineers to debug schema mismatches and data pipeline failures, reducing resolution time by 25%.
- Participated in exploratory **anomaly detection** and validation checks for real-time fraud and risk scoring pipelines.

Accenture Solutions Pvt. Ltd.   *Pune, India*
Test Analyst – Backend & ETL Validation   *Nov 2018 – Apr 2022*

- Performed **SQL-based backend validations** and data integrity checks across large-scale CRM systems and reporting workflows.
- Validated ETL pipelines and APIs supporting customer data processing and operational dashboards.
- Automated regression testing using Python and custom scripts, reducing manual QA cycles by 40%.
- Supported **ML pipeline integration testing** to ensure scalability and data consistency in scoring models and analytics tools.
- Led defect triage and root cause analysis for high-priority data issues, improving SLA compliance and system stability.

Projects

- **RAG-Based QA App (LLM)** — Built a Retrieval-Augmented Generation (RAG) application using Groq LLMs and Streamlit for document-based question answering. Implemented vector search and prompt design to improve answer relevance and reduce hallucinations during manual evaluation.
- **NER Token Classification** — Fine-tuned a BERT-based Named Entity Recognition model on annotated datasets, achieving F1-score of 0.85. Improved entity detection accuracy through iterative annotation refinement and hyperparameter tuning.
- **SaaS Data Warehouse Pipeline** — Developed a Python–SQL ETL pipeline with automated validation and anomaly detection, identifying data inconsistencies across multiple sources and reducing manual data inspection during test runs.
- **House Price Prediction** — Trained regression models (XGBoost, Random Forest) on 21k+ property listings, achieving R^2 of 0.80. Applied feature engineering and cross-validation to improve generalization and mitigate overfitting.
- **Neuron Morphology Optimisation (Dissertation)** — Built a scalable BioDynaMo-based simulation pipeline to optimise neuron structures, enabling large-scale parameter calibration across hundreds of simulation runs and improving convergence stability in experimental results.

Certifications

- Python for Data Science — LinkedIn Learning (Aug 2025)
- SQL for Data Science — LinkedIn Learning (Aug 2025)
- Microsoft Certified: Azure AI Fundamentals (May 2025)
- Postgraduate Program in Data Science and Business Analytics — Great Lakes Institute of Management (2023)

Additional Information

- Event coordination at university-level technical and cultural activities
- Volunteering: National Service Scheme — literacy & blood donation drives, rural literacy drives
- Active member, AI and Data Science Society – University of Surrey
- Volunteering: University of Surrey — community gardening and sustainability projects