

SOUMYAJIT ROY

P: +91 9777684766 | roysoumyajit@icloud.com

SUMMARY

Data Scientist with 1+ year of experience in fintech, working on credit analytics, transaction intelligence, and ML/NLP systems at scale. Strong experience in feature engineering, model evaluation, and deploying ML solutions on large transactional datasets to drive measurable improvements in decisioning quality, cost, and operational efficiency. Strong computer science foundation with hands-on experience translating data insights into production-ready solutions.

WORK EXPERIENCE

PINE LABS (Leading fintech company in Asia)

Bangalore, India

Data Scientist - I

Aug 2025 – Present

- Designed a recurring transactions detection module using **trigram Jaccard similarity** and **union-find clustering**, improving precision by **15%** and recall by **30%** over the legacy system, while reducing inference latency by **50%** and lowering annual infrastructure costs by **\$6,000**.
- Developed a **salary detection pipeline** from bank statements using **BERT-based NLP models**, improving recall by **20%** on real-world noisy financial data.
- Built **Logistic Regression** and **CatBoost** models for income estimation from alternate data, leveraging feature engineering and hyperparameter tuning to achieve **~75% accuracy** and improve downstream credit decisioning quality.
- Designed a **time-series forecasting system** for payment traffic by benchmarking **XGBoost**, quantile regression, and **Temporal Fusion Transformers (TFT)**; final solution reduced infrastructure provisioning costs by **~50%**.
- Applied **prompt engineering** with **Qwen Vision-Language Models (VLMs)** to extract structured data from PDF bank statements for downstream ML workflows.

PINE LABS (Leading fintech company in Asia)

Bangalore, India

Data Science Intern

Sep 2024 – Jul 2025

- Fine-tuned Qwen-2.5 LLM to automate annotation review process, reducing manual reviewer workload by **25%**.
- Improved heuristics-based models for retail underwriting and personal finance management (PFM) by analysing false positives and negatives on large-scale transaction data, increasing both precision and recall by **15%**.
- Processed and analysed **50M+ transaction records** to extract actionable insights, enabling successful proof-of-concept (POC) initiatives with internal teams and external stakeholders.

SKILLS

Programming Languages: Python, C, C++, Java

Technical Skills: Large Language Models (LLM), BERT, Prompt Engineering, LLM Fine-tuning (QLoRA, LoRA), Logistic Regression, CatBoost, Feature Engineering, Machine Learning, Pandas, NumPy, Data Analysis, SQL

Tools & Platforms: Git, Github, Gitlab, AWS S3, AWS Sagemaker

EDUCATION

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

Bhubaneswar, India

Bachelor of Technology, Computer Science and Engineering

2021-2025

Cumulative GPA: 9.50/10.0

PUBLICATIONS

CodeAnubad: Efficient Bangla-to-Python Code Generation

- Built a Bangla-to-Python code generation system for BLP-2025 Task 2 at AACL-IJCNLP by fine-tuning Gemma-2-9B-it with QLoRA and LoRA rank optimisation.
- Designed an iterative self-improvement pipeline that augmented scarce training data (74 samples → 226 samples) with verified model outputs, boosting dev pass rate, outperforming translation and RAG baselines.

Unveiling Natural Based Underwater Image Colour Enhancement: a Paradigm Shift in Underwater Image Processing and Colour Enhancement

- Proposed the NUCE method using swarm intelligence techniques to significantly improve color distribution and visual quality in underwater imagery.

PlantDiseaseNet-RT50: A Fine-tuned ResNet50 Architecture for High-Accuracy Plant Disease Detection Beyond Standard CNNs

- Proposed PlantDiseaseNet-RT50, a fine-tuned ResNet50 architecture with selective layer unfreezing and cosine learning-rate scheduling, achieving ~98% accuracy/precision/recall with a lightweight deployment-ready model.

ACHIEVEMENTS

- **Mentee, Amazon ML Summer School 2023, 2024:** Got selected as one of the mentees in Amazon Machine Learning Summer School out of all candidates across India for two years in a row - 2023, 2024.
- Zonal Rank 14 and Gold Medal (for being school topper) in SOF National Science Olympiad 2019.