

ARYAN NAYAK

Data Engineering & ML Infrastructure

+91-891-723-7073 | aryannayak375@gmail.com | [linkedin.com/in/ary8n](https://www.linkedin.com/in/ary8n) | github.com/ary8n | ary8n.vercel.app

SUMMARY

Computer Science graduate (KIIT, 2026) with hands-on experience building and owning data pipelines end to end — ingestion, transformation, validation, and delivery — across financial and operational datasets at Infosys and PrimeTrade AI. Comfortable working from raw, messy data through to a shipped model: designed, trained, and deployed an independent recommendation system from a 6M-row public dataset, including the full ETL and evaluation pipeline. Trained on distributed data systems (PySpark, Kafka) through coursework and hands-on labs, and looking to apply that foundation to infrastructure that handles data at scale.

TECHNICAL SKILLS

- **Data Engineering:** Python, Pandas, NumPy, SQL, ETL Pipeline Design, Data Validation & Quality Control, MySQL, PostgreSQL, Stored Procedures, Query Optimisation
- **Distributed & Streaming Systems (coursework / hands-on labs):** PySpark, Apache Kafka
- **Machine Learning:** scikit-learn, XGBoost, LSTM, PyTorch, Predictive Modeling, Recommender Systems, Anomaly Detection
- **Analytics & BI:** Power BI (DAX, Power Query), Tableau (basic), Advanced Excel, KPI Reporting, Dashboard Development
- **Tools & Practices:** Git, GitHub, Jupyter Notebooks, Agile, Scrum

EXPERIENCE

Data Engineer Intern | Infosys

Dec 2025 – May 2026

- Owned an end-to-end reporting pipeline in Python and MySQL, applying validation and quality-control checks at each stage from ingestion through transformation to dashboard delivery.
- Wrote optimised SQL — stored procedures and multi-table joins — to extract and transform large structured datasets for operational reporting.
- Built Power BI dashboards on top of the data warehouse layer, translating pipeline output into KPIs for business stakeholders.
- Served as Scrum Master for a 5-member team across 2 delivery sprints while contributing as a hands-on developer.

Trader Behaviour Analysis | PrimeTrade AI

- Ran trend analysis and root-cause investigation across trader behaviour and sentiment signals (Fear/Greed/Neutral index) to identify drivers of trading outcomes.
- Built Power BI dashboards on portfolio-level KPIs (profit margin, win rate, drawdown) to surface performance patterns across trading strategies.

PROJECTS

Shelf Life — Recommender System from Raw Data to Model *(Independent)*

- Built, alone, an end-to-end recommendation pipeline on 5.98M real reader ratings and 912,705 'want to read' signals (goodbooks-10k): data cleaning, an item-item latent-factor model (truncated SVD), and a re-ranking layer.
- Engineered a 'follow-through rate' signal — the share of readers who shelved a book but never rated it — from raw behavioral logs, then blended it into the ranking to bias recommendations away from high-abandonment items.
- Shipped as an interactive, deployed data story with a live model behind it: [demo](#) | [notebook](#)

Machine Learning Prediction Pipeline

- Designed a predictive analytics pipeline (XGBoost + LSTM ensemble) achieving 92% forecasting accuracy on noisy real-world data, automated into modular, independently testable components.

Real-Time Vision Diagnostic System

- Built a real-time classification pipeline (PyTorch, ResNet18) at 90% accuracy, with entropy-based anomaly detection flagging low-confidence outputs — improving post-deployment accuracy by 5% via targeted retraining.

EDUCATION

B.Tech, Computer Science Engineering — Kalinga Institute of Industrial Technology (KIIT), Bhubaneswar

GPA: 8.51/10 | 2022 – 2026

Doon International School — Intermediate (12th, CBSE), 79%, 2022 | Little Flower School — Matriculation (10th, ICSE), 95%, 2020

CERTIFICATIONS

- IBM Data Science Professional Certificate | London Business School — Corporate Responsibility