RANJAN KUMAR CHOUBEY

Bengaluru ranjankumarchoubey999@gmail.com
LinkedIn GitHub Portfolio

Summary

Machine learning engineer with 3.5 years of experience taking models and data pipelines to production. Built and deployed ARIMA and XGBoost forecasting models at Tata Power, improved an XGBoost payment-default model at Allstate, and now build Spark-Scala data products on GCP for Walmart. M.Tech in Data Science, ISI Kolkata (AIR 11).

Skills

Machine Learning: scikit-learn, XGBoost, PyTorch, deep learning, feature engineering, time-series forecasting (ARIMA), statistics

LLM / GenAI: LLM APIs, RAG, FAISS, Hugging Face, fine-tuning (LoRA), AI agents (LangGraph), LangChain, LLM evaluation

Data: SQL, Apache Spark (Scala, PySpark), Airflow, BigQuery, Hudi, Kafka, Pandas, NumPy

MLOps & Engineering: Python, Scala, FastAPI, Docker, Kubernetes, MLflow, CI/CD, GCP (Vertex AI, GCS, Dataproc), Git, Linux

Experience

Consultant | ***Deloitte*** *May 2025 – Present | Bengaluru*

Client: Walmart International (since Jul 2025)

- **Enterprise Planning:** Independently delivered the Sales module – Spark-Scala pipelines that compute sales and margin over rolling N-week windows into Hudi and generate SAP BW planning feeds (6 feed types, 3 retail banners), validated with SAP and BW teams. Runs daily in about 30 minutes, feeding Walmart's daily planning.
- **Order substitutions:** Developed and released the substitution analytics layer end to end over Kafka order events (Spark-Scala, BigQuery) that powers nil/short-pick, substitution-outcome and fill-rate reporting.
- **Markdown reporting (current):** Leading discovery to deployment for a 7-site pilot – profiled SAP source data, defined 18 KPIs, and designed and validated 3 BigQuery views on markdown activity, approvals and waste.
- **Ingestion:** Brought 100+ SAP and Kafka source tables into the GCS datalake (Airflow, CI/CD).

Data Science Intern | ***Allstate India*** *May 2024 – Jul 2024 | Bengaluru*

- Improved an XGBoost payment-default prediction model through geographic feature engineering; presented findings to Bengaluru and US teams, leading to 3 model enhancements.

Lead Engineer (ICT) | ***Tata Power*** *Sep 2021 – Jul 2023 | Mumbai*

- **Generation forecasting (Haldia plant):** ARIMA and XGBoost models on real-time PI System sensor and Tata Steel data (temperature, pressure, boilers online, coal moisture), forecasting output in day-ahead time blocks.
- **Sales forecasting:** Models on IEX market data supporting day-ahead power trading.
- **Deployment:** Put both models into production on on-premise servers with CI/CD; created REST APIs (FastAPI, Node.js) serving 10+ Power BI dashboards.

Research & Projects

Reducing Attention Complexity in Graph Transformers through Subgraph Partitioning | ***M.Tech Dissertation, ISI Kolkata***

- Partitioned graphs into K subgraphs (METIS) and applied attention over a $K \times K$ supergraph of GCN + Laplacian-embedding tokens instead of the full $N \times N$ graph, cutting attention cost from $O(N^2)$ to $O(K^2)$. PyTorch Geometric. Supervisor: Swagatam Das.

Notebook App | ***Personal project***

- Shipped a full-stack note-taking app with LLM-assisted writing (Gemini, Groq APIs), code blocks, math rendering and PDF export. Next.js, TypeScript, Firebase.

Education

M.Tech, Computer Science (Data Science) | ***Indian Statistical Institute, Kolkata*** *2023 – 2025 | Kolkata*

Aggregate 78.5% | ISI entrance: AIR 11 (AIR 1, EWS category)

B.Tech, Information Technology | ***Birsa Institute of Technology, Sindri*** *2017 – 2021 | Dhanbad*

CGPA 8.89/10 | Rank 1 in department | GATE CSE 2021: AIR 1958

Teaching & Recognition

- **IMS Academy (part-time):** GATE DA faculty for Probability & Statistics and Machine Learning; subject-matter reviewer for the GATE DA 2027 mock-test series; designed a 30-problem DSA interview course.
- Recognised by the Walmart client manager for the Enterprise Planning and order-substitution deliveries (2026).
- Star of the Month (4x), Tata Power, 2022–2023.