

M. Ikram Ullah

Data Scientist · Experimentation, Causal Inference & Machine Learning

Islamabad, Pakistan | (+92) 311-7862537 | ikramullah4748@gmail.com | linkedin.com/in/ikram-ullah11 | github.com/ikram4748 | ikram4748.github.io

SUMMARY

Data scientist who turns experiments and data into decisions with honest uncertainty. Recent work: a full readout of a 64,000-customer randomized e-mail experiment (validity checks, Holm-corrected effects, CUPED, power and peeking analysis, uplift models and an IPW-valued targeting policy), a text-to-SQL analyst agent benchmarked on BIRD, and a retrieval benchmark over 84 company filings. Strong in Python, SQL and statistics; experienced with dashboards for non-technical stakeholders; pursuing an MS in Data Science.

TECHNICAL SKILLS

Statistics & experimentation: A/B testing, sample-ratio mismatch, power analysis and MDE, CUPED and regression adjustment, bootstrap, multiple-testing correction (Holm), sequential-testing pitfalls, causal inference

Machine learning: uplift modeling (S/T/X-learners, transformed outcome, Qini/AUUC), XGBoost, LightGBM, scikit-learn, time series, class-imbalance handling, SHAP/LIME, model monitoring

Data & SQL: advanced SQL (window functions, CTEs, query optimisation), DuckDB, SQLite, PostgreSQL, MySQL, MongoDB, pandas, NumPy, statsmodels, data pipelines

Visualisation & BI: Power BI (DAX), Tableau, Plotly, matplotlib, Streamlit dashboards

AI & engineering: LLM applications (text-to-SQL, RAG), LangGraph, Gemini/OpenAI APIs, FastAPI, Docker, Git, pytest

EXPERIENCE

Data Scientist / ML Engineer, Syslab Technologies (*Islamabad, Pakistan*) Oct 2024 – Nov 2025

- Improved diagnostic response accuracy by 23% through systematic A/B testing of model configurations, prompt strategies and RAG retrieval settings (chunk size, retrieval depth).
- Built an automated model-monitoring pipeline (accuracy, precision, recall, F1, latency) with real-time Power BI dashboards for early detection of model drift.
- Investigated model failures through error logs, edge cases and interaction data; targeted retraining and data-quality fixes reduced the error rate by 18%.
- Processed 50K+ patient interaction records with Python and SQL; validated outputs with product managers, QA and healthcare providers against clinical standards in an Agile team.

Freelance Data Scientist, Self-employed (direct clients) (*Remote*) Feb 2023 – Present

- Delivered 15+ end-to-end AI/ML projects for healthcare, e-commerce and fintech clients, from scoping and data collection to deployment and handoff.
- Built churn, demand-forecasting and fraud-detection pipelines (scikit-learn, XGBoost, LightGBM) with MLflow experiment tracking; reported results through Tableau and Power BI dashboards.

PROJECTS

Data Analyst Agent (text-to-SQL) (github.com/ikram4748/data-analyst-agent)

LangGraph · Gemini · sqlglot · DuckDB · fastembed · Streamlit

- Built a self-correcting text-to-SQL agent: local-embedding schema linking, a two-layer read-only SQL guard, a literal-value check that catches silent zero-result errors, and a bounded fix loop driven by database feedback.
- Raised execution accuracy on 150 held-out BIRD questions from 60.7% (one-shot baseline, same model) to 66.7% (+6.0 points, 95% CI +1.3 to +11.3) at 1.09 LLM calls per question; tuned on a separate dev split after error analysis of 48 failures.
- Found and fixed a measurement bug in which rate-limit fallback silently mixed three models across benchmark variants.

Financial RAG Lab (github.com/ikram4748/financial-rag-lab)

PyMuPDF · BM25 · bge-small · cross-encoders · Gemini

- Built a RAG benchmark on FinanceBench (150 analyst questions, 84 filings, 12,013 pages) comparing 3 chunking strategies, 3 metadata scopes and 5 retrievers (BM25, dense, hybrid RRF, two cross-encoder rerankers) plus HyDE, scored on evidence-page recall with a dev/test split.

- Measured end-to-end answer accuracy from closed-book to an evidence-page oracle, graded by a separate judge model validated on gold and corrupted answers.

Qanoon Assist: legal-information RAG for Pakistani law (private) *(code available on request)*

FastAPI · React · PostgreSQL · Qdrant · BGE-M3 · Gemini

- Scraped and parsed 6,664 federal and provincial laws (117,365 sections, 147,701 chunks) from five government portals; answers in English, Urdu and Roman Urdu with section-level citations.
- Hybrid BM25 + BGE-M3 dense retrieval with bge-reranker-v2-m3 and citation verification; hit@5 1.00, faithfulness 0.944 and 4.8 s average latency on a 36-question evaluation; 124 automated API and UI tests.

E-mail Experiment & Uplift Modeling *(github.com/ikram4748/uplift-ab-testing)*

Python · statsmodels · LightGBM · Plotly · Streamlit

- Analysed a 64,000-customer randomized e-mail experiment end to end: sample-ratio and balance checks, Holm-corrected effects (men's e-mail +7.7 points visits, +\$0.77 spend per customer, 95% CI \$0.49–\$1.05), CUPED and regression adjustment, power analysis, and a peeking simulation (10 looks → 18% false positives).
- Built S/T/X-learner and transformed-outcome uplift models with Qini evaluation over 5 splits and valued targeting policies by inverse-propensity weighting; showed that e-mailing everyone beats model targeting below about \$0.23 per e-mail.

FraudShield: credit-card fraud detection

XGBoost · LightGBM · SMOTE · SHAP · LIME

- Fraud detection on 50K transactions with a 1.7% fraud rate; SMOTE/SMOTETomek resampling, five models with threshold optimisation (0.95+ AUC-ROC), SHAP and LIME explanations for every prediction.

RESEARCH

UrduComicLab: complete-story Urdu story-to-comic generation (MS thesis, in progress)

- Open-weight pipeline turning complete Urdu stories into right-to-left comics; vs baseline: spoken lines printed 0.49 → 1.00, dialogue F1 0.63 → 1.00, face consistency 0.54 → 0.67 (paired bootstrap).

Facial emotion recognition on imbalanced FER2013: CNN vs Vision Transformer (paper, pending publication)

- CNN (SE blocks) vs ViT-Base/16 on FER2013 with class-balanced focal loss; ViT avoids minority-class collapse (Disgust 64.9% vs 0%).

EDUCATION

MS Data Science, Quaid-e-Azam University, Islamabad

Feb 2025 – Present

Coursework: Machine Learning, Deep Learning, NLP, Big Data Analytics, Statistical Modeling, Data Visualization. Thesis: UrduComicLab.

CERTIFICATIONS

Machine Learning Specialization · Deep Learning Specialization · Natural Language Processing · Big Data Analytics