Utkarsh **Tripathi**

Data Engineer
Bengaluru, India
Email: utkarsh1999tripathi@gmail.com | Portfolio: utkarshtripathi.com | StackOverflow: inarticulatus | Linkedin: utkarshtripathi7

Experience

RiskPe (Insurtech Startup) **Jan 2025 - Present**
Data Engineer **Jaipur, Rajasthan**
Keywords: AWS EC2, PySpark, Data Pipeline, Power BI, Python, SQL, Prophet Forecasting

- Reduced **reporting latency by 40%** by architecting automated ETL pipelines processing 10K+ daily policy and lead records using PySpark.
- Improved **conversion tracking accuracy** by building Power BI dashboards surfacing underwriting KPIs and fraud indicators across 5 business units.
- Increased **sales forecasting precision by 15%** by developing Prophet-based time series models, directly informing quarterly planning.
- Designed end-to-end automated workflows for lead leakage and agent performance tracking, enabling leadership to take corrective action in real-time.
- Drove analytics roadmap definition in partnership with founding team during company pivot to AI-driven consultation flows.
- Shipped version-controlled SQL templates and data quality checks, reducing analyst rework by eliminating silent data errors.

Sheshsuki Automation Pvt. Ltd. (Habitat) **July 2024 - Dec 2024**
Computer Vision Engineer **Bangalore, Karnataka**
Keywords: YOLOv7, OpenCV, Edge Inference, Embedded Systems, Cost Optimization, R&D

- Reduced **ANPR system costs by 50%** by deploying YOLOv7-based solution on BeagleY-AI edge hardware, replacing expensive commercial systems.
- Achieved **real-time inference at 15+ FPS** with OpenCV optimizations, enabling scalable license plate detection on low-power ARM processors.
- Improved **detection accuracy by 20%** under challenging lighting by augmenting training data and tuning post-processing heuristics.
- Built modular edge inference pipeline enabling rapid A/B testing of YOLO weights and preprocessing configurations.
- Collaborated with hardware team to benchmark latency metrics across CPU cores, achieving measurable performance gains.

Siemens Ltd. **July 2021 - June 2023**
System Engineer **Bangalore, Karnataka**
Keywords: Python, Power BI, Engineering Analytics, ETL, Resource Forecasting

- Built **Python + Plotly & Power BI pipelines** serving analytics for 3 global rail projects: Vienna Metro, Sydney Airport Line, and Indian Railways.
- Reduced **manual data processing time by 30%** by automating condition monitoring ETL scripts for rolling stock diagnostics.
- Developed failure prediction scripts enabling proactive maintenance scheduling and early warnings for critical components.
- Consolidated **5+ years of sensor data** into standardized analytical datasets, accelerating engineering investigations.
- Coordinated with cross-functional teams across Europe and India to translate engineering requirements into visual analytics tools.
- Created reusable Plotly visualization library, reducing build time for new reporting modules by 40%.

Education

Diploma, Data Science & Applications, Indian Institute of Technology Madras, Chennai, Tamil Nadu **2025**
BSc, Transportation Technology, National Rail and Transportation Institute, Vadodara, Gujarat **2021**

Skills

Core: Python · SQL · Apache Spark · Kafka · Flink · Airflow
Cloud & Infrastructure: AWS (S3, Redshift, Glue) · Docker · Kubernetes · Terraform · Git & CI/CD
Data & Analytics: Data Modeling · dbt · Vector DBs (Qdrant) · Redis · Power BI
AI & LLM: LangChain · Sentence-Transformers · FastAPI · RAGAS Evaluation

Projects

Real-time E-commerce Recommender System **2025**
Keywords: Kafka, Flink, Spark, Airflow, Vector DB, MLOps

- Architected Lambda Architecture pipeline processing **1M+ daily clickstreams** via Kafka, Flink, and Spark.
- Implemented Two-Tower Retrieval & DeepFM Ranking system, serving **1M+ item embeddings** with sub-100ms latency via Milvus.
- Orchestrated **Airflow DAGs** for daily feature computation, weekly model retraining, and drift detection alerting.

Production RAG Pipeline for Document Intelligence **2025**
Keywords: LangChain, Qdrant, dbt, Sentence-Transformers, FastAPI

- Built end-to-end RAG system ingesting **10K+ documents** into Qdrant with hybrid retrieval (dense + BM25).
- Developed **dbt transformation layer** for document metadata, chunk statistics, and evaluation metrics.
- Improved **answer precision by 25%** by implementing cross-encoder reranking and RAGAS evaluation.

ML From Scratch: Gradient Descent to Gradient Boosting **2025**
Keywords: NumPy, Optimization, Gradient Boosting, Algorithm Design

- Implemented **core ML algorithms** from first principles using NumPy: Linear/Logistic Regression, Decision Trees, Random Forest, GBM.
- Built XGBoost-style GBM with second-order Taylor expansion, achieving **<2% accuracy gap** vs sklearn on benchmark datasets.