ROUSHAN KUMAR

India • +91-8986510661 • roushankumarmail07@gmail.com • linkedin.com/in/rk0718 • github.com/rkuma18 • itsrkumar.com

PROFESSIONAL SUMMARY

Gen AI Engineer | 6+ YOE specializing in **LLM systems, RAG, fine-tuning**. Built RAG applications with **LangChain, GPT-4, vector databases**; fine-tuned **DistilBERT achieving 96%+ F1-score**; engineered ML pipelines processing **10M+ records/day**. Delivered record-breaking sales performance and 45% waste reduction through data-driven optimization.

SKILLS

Gen AI & LLMs:	LangChain, RAG (FAISS, ChromaDB), GPT-4, Transformers, Hugging Face, Fine-Tuning, Prompt Engineering
ML/DL:	PyTorch, Scikit-learn, XGBoost, NLP, Time Series Forecasting
MLOps & Cloud:	AWS (SageMaker, Bedrock, S3), Docker, MLflow, FastAPI, CI/CD
Data & DBs:	Python, SQL, PL/SQL, PostgreSQL, MySQL, R
Tools:	Tableau, Streamlit, Git, Jira

WORK EXPERIENCE

JD Wetherspoon **Apr 2023 – Feb 2025**
Operations Analyst - Business Analytics *London, UK*

- Built **SQL-based demand forecasting** analyzing 2+ years of sales data to optimize staff scheduling and operations, contributing to outlet achieving **record-breaking £55k weekly sales** (highest performance across all company locations).
- Developed **Python/SQL inventory optimization** reducing waste by **45%** through sales velocity analysis across 200+ SKUs to minimize spoilage and improve capital efficiency.
- Prototyped **RAG-based knowledge base** using **LangChain, ChromaDB** for internal documentation retrieval, demonstrating Gen AI applications in operational workflows.

Tata Consultancy Services **Mar 2019 – Jul 2021**
Data Analyst *Pune, India*

- Conducted **exploratory analysis (Python/SQL)** to drive strategic segmentation and product insights for BMO's personal & commercial banking products.
- Built real-time **BI dashboards (Tableau)** to monitor user engagement and product performance, enabling data-driven decision-making.
- Improved reporting accuracy by **20%** via data mining and validation, optimizing sprint planning and backlog visibility.

Tata Consultancy Services **Jun 2017 – Mar 2019**
Backend Developer *Pune, India*

- Engineered **ETL pipelines (Python, PL/SQL)** processing **10M+ records/day** from global servers to standardize data for downstream analytics and ML training datasets.
- Transformed complex raw **JSON feeds** into optimized normalized tables to accelerate ML model training and analytics applications.
- Reduced data ingestion errors from **5% to 1%** by optimizing SQL queries, leading code reviews, and automating validation checks.

PROJECTS

DocuChat AI - Document Intelligence System *[GitHub]*

- Built **RAG application** with **LangChain, FastAPI, ChromaDB** for multi-document Q&A; supports PDF/DOCX/HTML with history-aware retrieval and session persistence. **Tech:** Python, OpenAI, Streamlit.

Financial SMS NER - Entity Extraction *[GitHub]*

- Fine-tuned **DistilBERT on 50k SMS** extracting **20+ entities** achieving **96.2% F1-score**; optimized inference to **<20ms/batch** for production. **Tech:** PyTorch, Transformers, Hugging Face.

FX Intelligence Agent - Multi-Currency AI *[GitHub]*

- Built **GPT-4 agent with function calling** for currency conversion supporting **170+ fiat currencies and 14 cryptos**; deployed on Hugging Face Spaces. **Tech:** LangChain, OpenAI, Streamlit.

EDUCATION

MSc Data Science & Analytics, Cardiff University *2021–2023*
B.Tech Computer Science, SRM University *2013–2017*

CERTIFICATION

Retrieval Augmented Generation (RAG) *2025*
Generative AI with Large Language Models *2023*