Nitesh Kumar

Data Scientist - LLM Specialist, IDFC FIRST Bank | Ex Goldman Sachs | BITS Pilani

Senior ML Engineer specializing in Large Language Models (LLMs) with 4.5+ years of experience in building production-grade AI systems. Proven expertise in developing high-performance solutions including SQL query generation, image generation, and automated text generation platforms. Track record of implementing end-to-end MLOps pipelines and optimizing model performance through advanced techniques.

✉ nit.krsha@gmail.com

📱 +91 70051 56772

📍 Mumbai, India

in linkedin.com/in/nit-krsha

○ github.com/krsha-nit

EDUCATION

B.E. (Hons.)- Mechanical BITS Pilaini, Pilani (8.37 CGPA)

SKILLS

Generative AI: Prompt Engineering, RAG, agents, Finetuning, Quantization, Hugging Face, LangChain

LLMs used: GPT-4, 4o, DALL-E, GPT-4 vision, mistral, Llama, , Stable Diffusion, Imagen, gemini

DL/ML: TensorFlow, PyTorch, Regression, Classification, Clustering, RF, GBDT, XGBoost, SVM, PCA, RNN, CNN, Transformers

NLP: Spacy, NLTK, Text preprocessing, Sentiment analysis, Categorization, NER

MLOps: CI/CD, Docker, Kubernetes, Model Optimization, system design

Cloud/Databases: AWS (Sagemaker, Athena, S3), GCP Vertex AI, Azure, MongoDB, Redis Vector DB, Snowflake

Programming: Python, SQL

Tools: Git, streamlit, Alteryx, Tableau, Confluence, jira, Airflow

WORK EXPERIENCE

Prompt Engineer
IDFC FIRST Bank

10/2023 - Present   *Mumbai*

Achievements/Tasks

□ Architected an **LLM-powered SQL generation system** achieving **98.8%** accuracy and **5 sec** latency for complex queries, integrating Azure OpenAI with AWS Athena and Redis Vector DB (saving **5hrs**/query)

□ Developed a production-grade image generation platform using **DALL-E 3** and **GPT-4V**, reducing content creation time by **90%**.

□ Engineered an AI-powered marketing content generation system supporting multiple channels (SMS, Email, WhatsApp), reducing content creation time by **95%**.

□ Implemented model optimization techniques including **quantization** and **fine-tuning** of open-source models (**Llama 3, Stable Diffusion**).

□ Built comprehensive **MLOps pipelines** with automated testing and monitoring frameworks.

Data Scientist
Sears Global Technology India PVT. LTD.

05/2022 - 10/2023   *Pune*

Achievements/Tasks

□ Leveraged **LLM** models for **sentiment analysis** and summarization of customer reviews, streamlining analysis processes.

□ Developed an **automated email assistant** using **OpenAI** models for drafting and summarization.

□ Implemented multiclass classification for **product categorization**, improving inventory management.

□ Created financial and operational metrics **dashboards**, reducing manual work by **10 hours weekly**.

□ Analyzed profitability of backorder parts through comprehensive analysis and published a tableau dashboard along with **Forecasting** to modify targets accordingly.

Senior Data Analyst
Goldman Sachs

07/2020 - 04/2022   *Bengaluru*

Achievements/Tasks

□ Conducted dormancy analysis using **Random Forest classifier** for institutional clients, providing actionable retention strategies through data-driven insights.

□ Enhanced data exploration and analysis through Python for exploratory data analysis (EDA) and process automation. Developed **high-quality Tableau dashboards** to support decision- making and streamline operations.

□ Implemented **ETL** processes using **Alteryx** and **SQL** to automate data extraction along with daily report and team assignments saving **1.5 hours** of daily manual work.

□ Designed **SOPs** for defense design, ensuring robustness, and conducted **RCAs** for production issues related to productionized **Alteryx workflows** and **dashboards**.

Data Analyst Intern
Goldman Sachs

02/2020 - 06/2020   *Bengaluru*

Achievements/Tasks

□ Created an **email automation** tool using **VBA**, a macro-enabled Excel sheet, significantly reducing manual work by **18 hours per week** with eliminating any human inaccuracies.

□ Extracted relevant data points from Outlook emails using **RegEx** and stored them in Excel for efficient data retrieval saving **40 hours of manual work**.

CERTIFICATIONS

Coursera
1. Generative AI with Large Language Models , 2. Machine Learning Engineering for Production (MLOps) Specialization 3. Deep Learning Specialization, 4. Machine Learning Specialization

Udemy
1. Tableau Advanced: Master Tableau in Data Science, 2. Complete Google BiqQuery Masterclass: Beginner to Expert