Varun Kumar

[Paschim Vihar, New Delhi 110063] | +91 8920872073 | [iamvarunkumar23@gmail.com] |
linkedin.com/in/itsvarunkumar/
https://site.varunkumarportfolio.com/

Summary/Profile

Results-driven Data Scientist and MLOps professional with 3.7+ years of experience creating and deploying data-driven applications solving complex business problems in IT Security, Telecommunication, and Retail domains. Proven ability to manage the complete project lifecycle: creating, developing, testing, deploying (with CI/CD, Docker, Azure), and monitoring ML systems. Proficient in predictive modeling, anomaly detection, LLMs (like RAG and contextual analysis), data processing, and analysis. Experienced in Python automation, MLOps disciplines, project pipeline management, reporting actionable insights with Sigma, client handling, and articulating technical results into business KPIs.

Skills

- **Programming & Databases:** Python, SQL, Bash Scripting | MongoDB | Vector DBs (Pinecone, FAISS, ChromaDB)
- **Data Science & Machine Learning:** Predictive Modeling (Classification, Regression, Time Series Analysis, Survival Analysis - Cox Model), Anomaly Detection, Recommender Systems, Natural Language Processing (NLP), LLMs, Generative AI, Statistical Modeling, Feature Engineering, Data Wrangling, Data Analysis
- **ML Libraries & Frameworks:** Scikit-learn, Pandas, NumPy, PyTorch, TensorFlow/Keras, Transformers (HuggingFace), XGBoost, LightGBM, Statsmodels, SciPy, PyCaret, LlamaIndex, LangChain
- **MLOps & Deployment:** MLflow, Docker, CI/CD (GitHub Actions, Azure DevOps), Git/GitHub, Model Monitoring (Azure Monitor), Azure ML Services
- **Cloud & Big Data:** Azure (Azure Functions, Blob Storage, Azure SQL), AWS (S3), Snowflake, Snowpark (Mention Spark/Kafka from experience section if relevant to target role)
- **Visualization & BI:** Matplotlib, Seaborn, Plotly, Sigma
- **Web Frameworks & APIs:** Flask, Django, REST APIs, Streamlit
- **Collaboration:** JIRA, Confluence, Agile Methodologies

Experience

Bengaluru| May 2024 – Present

- Designed and implemented an end-to-end stream processing pipeline using Kafka, Spark Streaming, and PyTorch Autoencoders for real-time anomaly detection, achieving sub-second scoring latency.
- Deployed containerized (Docker) models on Azure Kubernetes Service (AKS), reducing mean-time-to-detect critical anomalies by 60% and false positives by 25%.
- Implemented automated model monitoring and retraining pipelines via Azure ML Services, reducing manual intervention by 90%.
- Built an advanced RAG system using LangChain, Pinecone, and Gemini Pro API, improving retrieved context relevance by 30% using hybrid search and Cohere Re-rank.
- Established rigorous RAG evaluation using RAGAS, reducing LLM hallucinations by 40% and improving task success rate by 25%.
- Developed an AI code review assistant by fine-tuning CodeLlama (Hugging Face Transformers) on custom data, improving issue detection accuracy by 25%.
- Integrated the code assistant into GitHub Actions, reducing manual code review time by an estimated 30%.
- Developed a multi-modal search engine (Text/Image/Video) using OpenAI CLIP and ChromaDB, achieving <150ms P95 search latency and improving text-to-image search relevance (NDCG@10) by 20%.

DATA SCIENTIST | **TELEPERFORMANCE** | Gurugram | June 2022 – May 2024

- Developed ML applications predicting customer repayment propensity, contributing to a 20% collection rate rise and a 40% call volume decrease.
- Designed and automated KPI reporting using Azure scheduler, providing performance insights.
- Developed a strategy identifying optimal customers to contact, resulting in a 27% rise in successful connections.
- Built an ML pipeline recommending "best time to call," increasing right party contacts by 27%.
- Implemented advanced sequence modeling (fine-tuned BERT) for NER and Relation Extraction from documents, achieving 92% F1-score (a 20% improvement over baseline).
- Engineered robust data preprocessing pipelines, reducing data cleaning time by 40% and automating information extraction workflow (Dockerized) decreasing manual review by 75%.

- Deployed extraction pipeline on Azure Functions, monitored with Azure Monitor, maintaining <500ms latency.
- Developed a hybrid e-commerce recommendation engine (Collaborative Filtering, Content-Based with Sentence-BERT, LightGBM ranker), improving predicted CTR by 18% and diversity by 15%.
- Engineered candidate generation using FAISS for item embeddings, retrieving 200 candidates from 1M+ items in <50ms.
- Deployed recommendation system as a REST API (Flask, Docker) with A/B testing framework.
- Constructed a mini MLOps platform on AKS integrating MLflow, DVC, Kubeflow Pipelines, and Seldon Core.
- Implemented CI/CD pipelines (Azure DevOps) for automated ML lifecycle management, reducing average model deployment time from 2 days to 3 hours.
- Integrated SHAP and LIME for automated explainability reports, improving model transparency.

TECHNICAL OPERATIONS ENGINEER | **MOBIKWIK** | Gurugram | April 2021 – Dec 2021

Projects

LLM-Powered Artwork Description Generator

- Developed a web application using Python (Django) and Google Gemini API to automatically generate marketing copy, tags, and hashtags for digital artwork based on user inputs.
- Implemented a user selection mechanism to improve efficiency and minimize manual edits.

Benchmark for LLMs

- Developed a benchmark to evaluate text generation performance across 10+ large language models, including Gemini, OpenAI, and Llama, using Python and Transformer libraries.
- Implemented a suite of evaluation metrics, including BLEU, ROUGE, and cosine similarity, to quantify the accuracy and relevance of generated text compared to a custom dataset of question-answer pairs.
- Automated the testing and evaluation process using an Agile methodology, completing the project within six sprints and reducing manual evaluation time by an estimated 40%.
- Deployed a web application on Firebase Studio to showcase the benchmark results, providing a user-friendly interface for visualizing model performance and facilitating easy comparison.

Significant Open Source Contribution (Scikit-learn)

- Identified and fixed a reported bug in Scikit-learn's IsolationForest implementation related to specific edge cases, improving model robustness.
- Wrote comprehensive unit tests achieving 100% code coverage for the modified module.
- Submitted changes via Pull Request and participated in code review, leading to the PR merge.

Education

M.SC IN DATA SCIENCE | **Liverpool John Moores** | Liverpool, UK | July 2022

POST GRADUATION DIPLOMA IN DATA SCIENCE | **IIIT-BANGALORE** | [Bangalore, India] | August 2021

B.TECH IN COMPUTER SCIENCE | **GITM, GURUGRAM** | [Gurugram, India] | June 2017