Shivam Johri

+91-9108404965 | shivamjohri247@gmail.com | linkedin.com/in/shivam-johri | github.com/Shivamjohri247

Summary

Highly accomplished Senior Machine Learning Engineer with 9+ years of experience (5+ in AI/ML), specializing in architecting scalable MLOps pipelines, deep learning, NLP, LLMs, and cloud-based AI solutions. Proven expertise in model lifecycle automation, data engineering, and cloud technologies.

Technical Skills

MLOps
CI/CD, Airflow, MLflow,Spacy,Langchain, Huggingface,vllm,Ray

ML/DL
TensorFlow, PyTorch, Scikit-learn, LLMs, Deep Learning, Recommendation Systems, NLP

Big Data/Devops
Apache Spark, Hadoop, Hive, SQL,Docker , Kubernetes

Programming
Python, SQL, Go, Bash, Git/GitHub

Cloud
Google Cloud (Vertex AI, BigQuery, GKE), AWS, MS Azure

Databases
MySQL, NoSQL (MongoDB, HBase)

Experience

Senior ML Engineer — EPAM Systems — 02/2024 - Present

- Architected an end-to-end MLOps pipeline for a financial Named Entity Recognition (NER) model using BERT Transformers, improving search relevancy and results accuracy by 44%.
- Developed LLM-based query rewriting for indexing services, increasing search relevance and accuracy.
- Built a robust recommendation engine for stock name prediction, integrating it into a real-time search platform for search enhancement.

ML Engineering Senior Analyst — Accenture — 09/2021 - 02/2024

- Led development of LLM-based query expansion for financial queries, increasing query coverage by 25%.
- Created Signature detection models for forgery matching, achieving a 91%accuracy rate in detecting handwritten signed documents(Object detection usecase).
- Engineered deployed 5 personalized fashion recommendation models on GCP, improving engagement by 15%.
- Automated regulatory report generation, reducing manual effort by 40%.

Machine Learning Engineer — Tata Consultancy Services — 03/2016 - 09/2021

- Developed an end-to-end text detection and extraction pipeline for ID cards using CTPN, automating data processing.
- Built classification models for financial document processing, reducing validation costs by 30%.
- Led ML model deployment on Azure for banking clients, streamlining cognitive business processes.
- Implemented a use case for Extraction and Indexing using email data along with EntityRecognition

Education

B.Tech in Electronics and Communication Engineering
APJ Abdul Kalam University, Lucknow (2011 - 2015) — GPA: 6.24/10

Certifications

Google Cloud Associate Cloud Engineer (Valid till 11/2025)