

ANKIT SAXENA

ankit282k@gmail.com | +91 7500335301
linkedin.com/in/ankit282k | github.com/ankit282k

PROFESSIONAL SUMMARY

Machine Learning and Data Engineer with 1.5+ years of experience building scalable data pipelines, production-grade ML systems, and GenAI applications. Strong hands-on expertise across PySpark, Databricks, SQL, Python, and modern ML/LLM frameworks. Proven ability to deliver end-to-end solutions spanning data engineering, machine learning, and retrieval-augmented generation (RAG) systems. AWS Certified Cloud Practitioner with experience deploying secure, cloud-ready analytics and AI workloads.

CORE SKILLS

Data Engineering

PySpark, Databricks, Delta Lake, Snowflake, SQL, ETL Pipelines, Data Modeling, Medallion Architecture, Kafka

Machine Learning & GenAI

Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), NLP, Embeddings, Semantic Search, CNN, LSTM, Time Series Forecasting, Recommender Systems

Frameworks & Tools

Python, PyTorch, TensorFlow, HuggingFace, LangChain, FastAPI, MLflow

Cloud & MLOps

AWS (S3, EC2, IAM, Redshift, CloudWatch), Azure AI Foundry, Docker, GitHub Actions, CI/CD, Model Monitoring

PROFESSIONAL EXPERIENCE

Associate Software Developer — Bosch Global Software Technologies

Feb 2025 – Present

- Developed and deployed **LLM-based microservices** and data-driven automation solutions, improving team productivity by **32%**.
- Built scalable **NLP and embedding pipelines** using PyTorch, HuggingFace, and TensorFlow, reducing inference latency by **28%**.
- Designed reusable AI components enabling a **multi-use architecture**, reducing development time for new use cases by **40%**.
- Implemented production-grade applications using **Python, Spark, FastAPI, Docker**, with monitoring for model drift and performance regression.
- Collaborated with cross-functional teams in **Agile/SAFe** environments to translate requirements into reliable AI and data solutions.

Data Engineer Trainee — FranConnect Pvt. Ltd.

Sep 2024 – Feb 2025

- Engineered scalable **ETL and feature pipelines** using PySpark, Hive, and SQL, improving processing performance by **45%**.
- Optimized ingestion of **30M+ records** across distributed systems, ensuring data quality, governance, and lineage.
- Reduced Spark job costs by **22%** through partitioning, caching, and schema evolution strategies.
- Automated data and ML pipeline deployments using **Docker and GitHub Actions**, reducing deployment failures by **80%**.

Data Scientist Trainee — Dyno India Pvt. Ltd.

Feb 2023 – Aug 2023

- Built and evaluated ML models (Logistic Regression, Random Forest, XGBoost, SVM, K-Means, PCA, ARIMA) achieving **84–92% accuracy**.
- Developed deep learning models (**CNN, LSTM**) for text and image data, reducing validation loss by **30%**.
- Implemented recommender systems using Collaborative Filtering and FISM, improving precision by **25%**.

PROJECTS

Brain Box – RAG-based GenAI Chatbot

- Built an end-to-end **Retrieval-Augmented Generation (RAG)** chatbot over private documents.
- Implemented ingestion, chunking, and embedding pipelines using **LangChain**, storing vectors in **ChromaDB**.
- Integrated **Azure AI Foundry** models for secure LLM inference and enterprise-ready deployment.
- Designed prompts and retrieval strategies to improve answer relevance and reduce hallucinations.

Real-Time Streaming & ETL Pipelines

- Built batch and streaming pipelines using **PySpark**, **Kafka**, **Databricks**, and cloud object storage.
- Implemented **Bronze → Silver → Gold** architectures to improve data quality and performance.

ML & Forecasting Projects

- Developed **CNN-based image classifiers** achieving **94% accuracy**.
- Improved demand forecasting accuracy by **21% MAPE** using **ARIMA** and **XGBoost**.

CERTIFICATIONS

- **AWS Certified Cloud Practitioner (CLF-C02)** – Amazon Web Services, 2025

EDUCATION

B.Tech – Electronics & Communication Engineering

Chandigarh Engineering College | 2023 | CGPA: 7.93