

Mohd Arif Ansari

AI & ML Engineer

AI & ML Engineer with 4+ years of experience delivering **intelligent solutions** across **cloud platforms** using **machine learning**, **Big data**, and **real-time streaming technologies**. Skilled in developing **end-to-end ML pipelines**, including **data preprocessing**, **model training**, and **deployment**, as well as designing **cloud-native AI architectures**. **Certified** in **Azure AI**, **Data Engineering**, and **Data Science**. Demonstrated ability to **collaborate cross-functionally** and effectively **translate business challenges** into **scalable AI models** and **data-driven solutions**.



ariflpu786@gmail.com



+91-9779383344



Gurugram



linkedin.com/in/mohd-arif-ansari-43b24712b



github.com/Arif0000

SKILLS

AI/ML Technologies:-
Scikit-learn, PyTorch,
TensorFlow, MLflow,
Azure ML Studio, Model
Deployment .

Data Engineering for ML:
Spark, Kafka, Azure Data
Factory, Delta Lake,
Feature Engineering.

Cloud & Platforms: Azure
(Databricks, Synapse,
Blob Storage, ADF,
Cosmos DB), Docker,
GitHub Actions.

Programming: Python
(Pandas, NumPy,
PySpark), SQL.

Data Analytics &
Visualization: Power BI,
SQL Server, PostgreSQL,
MongoDB

LANGUAGES

English
Full Professional Proficiency

Hindi
Full Professional Proficiency

EDUCATION

Postgraduate in Advance Data Science and ML IIT Madras

03/2023 - 03/2024

Bachelor of Technology in Computer Science Lovely Professional University

07/2016 - 07/2020

EXPERIENCE

AI & ML Engineer Ingenious e-brain Solution

01/2023 - P

Outcome

- Built end-to-end ML pipelines using **PySpark** for scalable data preprocessing and feature engineering, processing over **10 million records** weekly.
- Developed **real-time streaming applications** with **Apache Spark and Kafka** in **Azure Databricks**, reducing data ingestion latency by **40%** and supporting **5K+ events/min**.
- Implemented intelligent document parsing using **NLP (Spacy, Tesseract)** to extract key fields from unstructured PDFs, improving document processing accuracy by **85%**.
- Designed and automated **web scraping pipelines** to collect training data from 100+ sources, accelerating dataset creation time by **70%**.

AI/Data Analyst Provana India

09/2023 - 12/2024

Outcome

- Architected a scalable Azure-based deployment pipeline for ML models handling **retail stream data**, enabling **real-time predictions** with **< 500ms latency**.
- Engineered PySpark pipelines to process **large-scale IoT data (500GB/day)**, improving feature generation time by **35%**.
- Built and deployed anomaly detection models in **Azure Synapse**, achieving **92% model accuracy** and detecting **geo-location fraud** across **50K+ devices/day**.
- Automated ingestion and retraining workflows using **Azure Data Factory and Event Hub**, reducing model update time from **weekly to daily**.

CERTIFICATES

Microsoft Certified: Azure Data Scientist Associate

Microsoft Certified: Azure Data Engineer Associate

Microsoft Certified: Data Analyst Associate