ARPIT PANDEY

DATA SCIENTIST

Phone +91-8359808703 | **Email:** arpitpandey6599@gmail.com | **LinkedIn** | **Github**

SUMMARY

Artificial Intelligence (AI) and Analytics Specialist with expertise in predictive modeling, data pipeline optimization, and advanced algorithm development. Proficient in leveraging cloud-native tools, supervised and unsupervised learning techniques, and generative AI for innovative solutions. Skilled in process automation, exploratory data analysis, and AI-driven strategies to deliver measurable business outcomes.

EDUCATION

- **Indian Institute of Technology Delhi**
  Post Gratuate Certification (Jun 2024-Jan 2025)
- **University Institute of Technology RGPV, Bhopal**
  B.Tech. in Information Technology (Jul 2018 - Jun 2022)

WORK EXPERIENCE

ARMAKUNI | AI/ML Engineer | Ahmedabad — Mar 2025 - Present

- Contributed to the development of Call Assist Intelligence, a GenAI-driven voice assistance platform utilizing AWS **Bedrock**, **Langfuse**, **Cloud Sonnet**, and Amazon Transcribe for speech-to-text conversion and model observability.
- Developed FastAPI-based backend services to integrate LLMs and transcription pipelines, enabling seamless voice interaction workflows and improving response accuracy and latency.

FISERV | Professional Data Analysis I Pune — Aug 2022 - Mar 2025

- **Fraud Detection Model**: Developed and deployed a real-time Fraud Detection Model using an **XGBoost**-based classification model with F1 score of 91%, integrated via FastAPI and Docker on **AWS** ECS achieving 30ms response time.
- **Fraud Alert Recommendation Model:** Built a **Random Forest** model for fraud alert decision recommender system, achieving an F1 score of 88%, leveraging advanced data preprocessing and statistical analysis techniques.
- **MLOps Framework**: Architected a scalable MLOps framework for automated model retraining and optimization, utilizing AWS SageMaker, AWS (S3, Lambda, Glue), Snowflake, GitLab (CI/CD Pipeline) to ensure seamless workflows.
- **Entity Matching and GenAI**: Designed a vector embedding based entity-matching system and implemented a Generative AI proof of concept using Amazon Bedrock to summarize fraud alert details with LLMs.
- **ETL Pipeline**: Developed and managed end-to-end ETL pipelines using Talend for integrating diverse data formats into a Snowflake data warehouse, ensuring reliability and scalability.

PERSISTENT | Python Intern | Nagpur — Jan 2022 - Jul 2022

- Acquired proficiency in **Python**, encompassing essential concepts and libraries such as Numpy, Pandas, Matplotlib, Scikit-learn and Neural Networks for implementing data analytics and machine learning tasks.

TOOLS AND FRAMEWORKS

- Rest APIs (FastAPI)
- Amazon Bedrock
- Big Data Platform (Snowflake)
- AWS (S3, EC2, ECS, Lambda, Glue)
- MLFlow
- Docker
- Amazon Sagemaker
- ETL (Talend)
- Git/GitLab (CI/CD Pipeline)

SKILLS

- **Languages and Libraries**: Python, SQL, Linux, Pandas, Polars, Scikit-Learn, Tensorflow, Keras, Pytorch, OpenCV.
- **Data Analytics**: Data Preprocessing, Data Analysis and Visualization, Dimentionality Reduction (PCA), Feature Engineering and EDA, Statistical Analysis, Parameter Tuning, Error Analysis (Precision/Recall/ F1 Score).
- **Machine Learning (ML)**: Linear Regression, Logistic Regression, Classification (Decision Tree, SVM), Unsupervised Learning (Clustering), Ensemble Learning (Bagging, Boosting, AdaBoost), Predictive Modeling
- **Deep Learning**: Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), LSTM, GRU, Computer Vision, Natural Language Processing (NLP).
- **Generative AI (GenAI)**: Transformers, Diffusion Models, Generative Adversial Networks (GAN), Retrieval-Augmented Generation (RAG), Large Language Models (LLMs).

PUBLICATION

- **An ordinal approach to Pneumonia detection from X-ray images using Convolutional Neural Network**

CERTIFICATION

- Machine Learning and Data Science with Python
- Deep Learning A-Z (Udemy)
- AWS Basics (Udemy)