

Shivam Choudhary

Data Scientist – Fractal Analytics



✉ Shivamchoudhary2014@gmail.com

📍 Hyderabad

☎ +91 9062330628

🌐 <https://linkedin.com/in/choudhary-shivam>

🔗 <https://github.com/shiv-choudhary>

🌐 <https://shiv-choudhary.github.io/portfolio>

Summary and Experience:

- Data Scientist specializing in **NLP, NLG & NLU** with **5+** years of experience developing Data Science solutions, Classical Machine Learning pipelines, Big Data Pipelines, Cloud Migration and REST API Development.
- Extensive domain experience in Media & Entertainment, Healthcare & Life Sciences, Insurance and Biopharmaceuticals.
- Expertise in deriving actionable insights from structured & unstructured data using Large Language Models (**LLMs**), Deep Learning (**DL**), Machine Learning (**ML**), **Statistics** techniques and **SQL**, particularly in **Product Management**.
- Proficient in designing **data-driven** solutions utilizing advanced technologies of unsupervised learning i.e. Topic Modelling using BERTopic, Transformer, Foundation and Hugging Face Models.
- Skilled in text classification, text-to-text generation, text summarization, conversational AI, Prompt Engineering, Retrieval Augmented Generation (**RAG**), **Multi Agentic Framework**, **Quantization**, **PEFT**, **LoRA**, **SFT**.
- Designed **reusable** components and **frameworks** employed by over 30+ teams to orchestrate AI pipelines on AWS.
- Designed a product analytics framework that **reduced call center RCR** from 26% to 8%, enabling the product manager to make informed decisions and assisting the manufacturing team in prioritization.
- Experience in Agile, Waterfall and Behavior Driven Development (BDD) approach of delivery.
- Received **YODA award** for creating a reusable analytics framework, which reduced the RCR, improved SLA & CSAT, while achieving substantial cost savings across call centers.
- Received the **Impact Award** for contributions to the design of an end-to-end pipeline capable of scaling inference to ~40 million records.
- Received **INSTA Award** for implementing a cost-effective data pipeline to process millions of unstructured data (PDF).

Education:

- Master of Computer Application (**MCA**) from the University of Mysore.
- Bachelor of Computer Application (**BCA**) from the West Bengal University of Technology.

Tools & Skillset:

- | | |
|---------------------|-------------------------------------------------------|
| 1. OS: | Linux, Win |
| 2. Language: | Python |
| 3. Database: | MySQL, MongoDB, Snowflake |
| 4. Cloud: | AWS, Azure, GCP |
| 5. Framework: | Tensorflow, Pytorch, Flask, FastAPI, GraphQL, PyUnit. |
| 6. GenAI Framework: | Langchain, LlamaIndex, LangGraph, CrewAI |
| 7. Tools: | Airflow, RabbitMQ |

Certifications:

- Microsoft Certified Azure Data Scientist (DP:100)
- Microsoft Certified Data Fundamental (DP-900)
- Microsoft AI Fundamental (AI-900)
- Microsoft Azure Fundamental (AZ-900)
- Microsoft Certified AI-Engineer (AI-100)
- Databricks certified Generative AI Engineer
- AWS Cloud Practitioner

Project Experience:

Fractal Analytics – (July 2023 – till date)

1. Hardware Device Insights

- ✓ **Description:** Designed a framework to analyze high-volume **unstructured** chat and call transcripts, deriving insights on customer satisfaction rates, repeat contact rates and SLA adherence for the Customer Satisfaction Index. Automated approximately 80% of the process, enabling efficient generation of actionable insights that informed stakeholder decisions and enhanced customer experiences. Utilized **classical ML techniques, transformer** models, **LLMs, SLM**, prompt definition and OOP concepts to ensure framework reusability.
- ✓ **Domain:** Entertainment & Media
- ✓ **Role Played:** Data Scientist
- ✓ **Contribution:**
 - i. Leveraged Unsupervised techniques like Clustering, HDBSCAN, UMAP and BERTopic.
 - ii. Leveraged LLM and Statistical techniques to support business-driven decisions.
 - iii. Utilized SLM(Gemma) and LLM(Gemini Flash 1.5) to understand customer behavior toward products, CSAT of unsubmitted customer surveys and agent behaviors.
- ✓ **Technology Used:** Python, LLM(Gemini Flash 1.5), BERTopic, Data warehouse, Topic Modelling, SQL

2. Advance Asia Analytics Platform

- ✓ **Description:** Designed, implemented & orchestrated AWS Cloud native solution for three use cases: generation of customer 360 data/view, smoker propensity use case and Top Advisor analytics. Combining three processed ~90 million data over the cloud. Migrated SAS ML pipeline to corresponding AWS services.
- ✓ **Domain:** Insurance
- ✓ **Role Played:** Data Scientist
- ✓ **Technology Used:** Python, AWS Sagemaker, Lambda, Step Function, DL, SQL

Carelon Global Solutions – (Sept 2022 – July 2023)

1. Care Management

- ✓ **Description:** Engineering & Orchestration of big data pipeline and ML models enabling Care Management applications for model inference. Developed a patient outreach analytics automation that will optimize the expenses of the organization and resource management.
- ✓ **Domain:** Healthcare
- ✓ **Role Played:** AI Engineer
- ✓ **Contribution:**
 - Created a pipeline to get data from Hive tables and mongoDB to perform the pre-processing feature using **Pyspark**.
 - Plugged **XGBoost-based** AutoML model in **BigData** Pipeline which will infer ~40 Million datapoints.
 - Implemented **SHAP** to understand local/global feature interpretation and feature importance while real-time inferencing.
 - Created a pipeline to summarize the clinical data for the ease of Nurses.
 - **Technology Used:** Airflow, Snowpark, MongoDB, PySpark, Hadoop, Yarn, Responsible AI, SQL

Deloitte USI – (Jan 2021- Sept 2022)

1. Data Migration and Analytics Platform

- ✓ **Description:** Migration of Big Data Stack, Reporting layer to Snowflake and exposed via restful service.
- ✓ **Domain:** Healthcare
- ✓ **Role Played:** AI Engineer
- ✓ **Contribution:**
 - Implemented cloud native framework for m
 - Migrated on-premises big data stack to Snowflake leveraging AWS data services (AWS glue, step functions)
 - Analyzing the big data stack and suggesting similar cost-cutting options to clients to optimize the cost of deliverables.
- ✓ **Technology Used:** Snowflake, Spark, Scala, AWS, SQL

2. Data Migration and Application Modernization

- ✓ **Description:** Application migration from Mainframe Technology to Digital Cloud Service (AWS) in the form of scalable microservices.
- ✓ **Domain:** Healthcare
- ✓ **Role Played:** Software Engineer
- ✓ **Contribution:**
 - Identified the layer of data and designed separate microservices for all the data access layer
 - Deployed designed microservices on EKS.
 - Written 100+ transformation rules to achieve graphQL user requests.
 - Integrated restful services with graphQL framework to reduce user requests' latency by 60%.
- ✓ **Technology Used:** Python, AWS, MongoDB, Flask-restful, graphQL, pytest, Docker, Kubernetes, SQL

Infosys Ltd - *(Sept 2019- Dec 2020)*

1. Data Remediation Tool

- ✓ **Description:** An in-house ETL workbench tool customization for the client, digitization of legacy paper recipes for modern MES system to be migrated and implemented.
- ✓ **Domain:** Biopharmaceuticals
- ✓ **Role Played:** Data Science Associate
- ✓ **Contribution:**
 - Designed a data pipeline that will extract data from PDFs and word files and dump it into PostgreSQL for data processing and data wrangling
 - Implemented ML model to generate keywords for building block suggestion and Template recommendation (BoW, TF-IDF, rake-NLTK, NER)