Shivam Choudhary
Software Engineer III – Carelon

✉ Shivamchoudhary2014@gmail.com   📍 Hyderabad

📞 +91 9062330628   in https://linkedin.com/in/choudhary-shivam

⎇ https://github.com/shiv-choudhary   🌐 https://shiv-choudhary.github.io/portfolio

Summary and Experience:

- Shivam is a Software Engineer III in Data Science capability having a total experience of ~**4** years, working primarily on the development of **Data Science** solutions, **Machine Learning** pipelines, **Big Data** Pipelines, **Cloud** Migration and **API** Development.
- He has engagement & domain experience with **healthcare** and life science clients (3 years), **Biopharmaceuticals** (1 Year).
- Evaluates the requirement and help with the options to make cost-optimized solutions for the benefit of clients and the organization.
- Expertise in designing data-driven solutions by leveraging cutting-edge technologies like **Pyspark**, **Airflow**, **AI**, MongoDB, cloud native, cloud agnostic solutions.
- Experienced in both **agile** and **waterfall** methodologies of delivery.
- Won INSTA award for implementing a cost-effective data pipeline that will process millions of unstructured data(pdf).
- Won Impact Award for the active contribution to designing an ETL pipeline to be integrated with ML models to scale the inference ~40 million record.
- **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
5. Framework: Flask, FastAPI, GraphQL, PyUnit, PySpark
6. 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)
- Microsoft Certified Python Programmer (MTA 98-381)
- AWS Cloud Practitioner

Project Experience:

Carelon Global Solutions – *(Sept 2022 – till date)*
1. Care Management
- ✓ **Description**: Engineering & Orchestration of big data pipeline and ML models enabling Care Management applications for model inference. Developed a patient outreach automation that will indirectly reduce the expenses of the organization and time.
- ✓ **Domain**: Healthcare
- ✓ **Role Played**: AI Engineer
- ✓ **Contribution**:
  - o Created a pipeline to get data from Hive tables and mongoDB to perform the pre-processing feature using **Pyspark**.
  - o Plugged **XGBoost-based** AutoML model in **BigData** Pipeline which will infer ~40 Million datapoints.
  - o Implemented **SHAP** to understand local/global feature interpretation and feature importance while real-time inferencing.
  - o **Technology Used**: Airflow, Snowpark, MongoDB, PySpark, Hadoop, Yarn, Responsible AI

Deloitte USI – *(Jan 2021- Sept 2022)*
1. Data Migration and Analytics Platform
- ✓ **Description**: Migration applications of Big Data Stack to Snowflake and exposed via restful service.
- ✓ **Domain**: Healthcare
- ✓ **Role Played**: Data Engineer
- ✓ **Contribution**:
  - o Migrated on-premises big data stack to snowflake leveraging AWS data services (AWS glue, step functions)
  - o Analyzing the big data stack and suggesting similar cost-cutting options to clients to optimize the cost of deliverables.
  - o Handling the project deliverables from offshore, plan Scrum planning to avoid the risk of delivery on time.
- ✓ **Technology Used**: Snowflake, Spark, Scala, AWS

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**: Data Science cum Data Engineer
- ✓ **Contribution**:
  - o Identified the layer of data and designed separate microservice for all the data access layer
  - o Deployed designed microservices on EKS.
  - o Written 100+ transformation rules to achieve graphQL user requests.
  - o Integrated restful services with graphQL framework to reduce user requests' latency by 60%.
- ✓ **Technology Used**: Python, AWS, MongoDB, Flask-restful, graphQL, pytest, Docker, Kubernetes

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**:
  - o Designed a data pipeline that will extract data from PDFs and word files and dump it into PostgreSQL for data processing and data wrangling
  - o Implemented ML model to generate keywords for building block suggestion and Template recommendation (BoW, TF-IDF, rake-NLTK)