

Gaurav Adlakha

gauravadlakha1509@gmail.com | +91 9717-033-977 | Bangalore, 560036, India

SUMMARY

- Data Scientist with more than 8 years of experience in machine learning and deep learning and 15 years of overall software development experience focussing mainly on Statistics, Econometrics and Machine Learning. Experienced in working with teams to integrate data science and analytics into decision making and portfolio of products and services.
- Experienced in implementing retrieval-augmented generation (RAG) systems and configuring vector databases for production applications.
- Designed and implemented serverless ML pipelines using AWS Lambda, Step Functions, and MLflow for experiment tracking, enabling scalable model training and deployment workflows
- Automated by building workflows to extract data from various REST APIs and databases, processing responses, data transformations in python
- Experienced in extracting and analysing data using SQL, web scraping techniques (Beautiful Soup, Scrapy) and API/web automation tools.
- Experienced in managing and analysing large datasets using various database such as SQLite, PostgreSQL, AWS, GCP, S3, HDFS, JSON, Hadoop and Parquet file formats.

PROGRAMMING LANGUAGES: Python, C++, Java

TOOLS & DATABASES: Spark, AWS, SAS, Hadoop, JFrog, Docker, Airflow, MongoDB, Minio, S3, EC2, Lambda, Kubeflow, Databricks, Argo, Plotly, Dash, Tableau, MS Visual Studio Code, Excel, Abbyy, OpenCV, Linux, Computer Vision, Tensor flow, Kubernetes, Vector Databases (FAISS, Pinecone, snowflake)

PROFESSIONAL EXPERIENCE

POWERSCHOOL- *Principal Data Scientist*

April 2022 –Present

- Developed a generative AI system to autonomously create content based on specified topics, grades, and educational standards, deploying the solution on a scalable serverless architecture.
- Created a chatbot that functions as a teacher's assistant, guiding students through problem-solving steps with hints and tips instead of direct answers, fostering independent learning and critical thinking skills.
- Created a natural language processing (NLP) model that translates manual language queries into SQL, querying databases set on snowflake, and presenting results in a summarized, user-friendly format, with the capability to engage in conversational output based on user queries.
- Implemented model safety guardrails leveraging open-source solutions and fine-tuned **Llama** language models.
- MLOps and ML product Development Roadmap ownership in collaboration with multiple product teams.
- Architected **scalable RAG pipelines** integrating vector databases to improve educational content retrieval accuracy.
- Implemented MLflow for experiment tracking and model registry, reducing model deployment time.

PWC-LABS- *Data Scientist, Advisory Division*

Sep. 2020 –April 2022

- Lead development of the firm's NLP ecosystem for information extraction, including a Python framework, open datasets, and a model hub. The project primarily involves the use of key environments such as Kubeflow, Azure, Linux, Python and OCR. It provides an extensible API for JSON I/O, text featurization, and model training based on **no-code configuration files** Using Kubeflow pipeline which can be used for batch inferencing, or batch execution, pipeline orchestrator.
- Leveraged Kubernetes for orchestrating distributed ML training jobs across multiple nodes.
- Planned and led delivery of skill recommendation engine using prerequisite learning for internal use, with in-year cost saving **3x for the firm.**

MPHASIS- *Data Scientist*

Mar. 2019 - Sep. 2020

- **Shipment Code Classification (more than 1000 categories)**-Worked on multiclass classification projects for FedEx where the responsibilities included.
 - Developing and implementing machine learning algorithms for large scale applications.
 - Collaborating with the client for an end to end design.
 - Utilized python to implement unsupervised machine learning document clustering technique for **1TB of unstructured data.**

- **Customer Invoice Information Extraction (Image regression, Image Segmentation, Information Retrieval)**-Designed and deployed a deep learning model for information extraction from customer invoices based on understanding 2D document, **which reduced the total process time by 20%.**
- **Cage Prediction**-Designed and deployed the machine learning model for cage prediction of shipment based on a entity embeddings of categorical variables to allow **50+ end client to prepare the shipment effectively.**
- Created infrastructure project to create and manage projects on AWS with Lambda using serverless architecture. Creation was simplified to One-Click and the cost for the production environment was lowered **from \$12000/month to \$1300/month an 89% cost reduction.**

ACCENTURE-Data Scientist

OCT. 2016- JAN. 2019

Worked closely with Business Analyst to develop data processing pipelines which handled data cleaning, error detection, statistical analysis and visualization with integration to external software.

- Developed Python scripts to parse, wrangle and analyze geophysics data.
- Designed machine learning algorithms on seismic data processing for Shell which performed feature extraction and tested deep learning algorithms such as image classification and segmentation for geological structure (faults, channels and salt boundaries) recognition.
- Created and managed data pipelines leveraging reusable patterns to expedite developer productivity such as **fastAI**.
- Developed application to generate text data from images and extract out required information from documents.
- Built machine learning (unsupervised) clustering python models (Hierarchical Clustering).

NUCLEUS SOFTWARE-System Analyst

FEB2014- OCT2016

- Project: Finn-one (Lending Product)
- Developed REST APIs for Field Investigation Module, a sub-system for Field Investigators, responsible for complete Field Investigation Workflow.

CSG- Senior Development Consultant

JUL2013-DEC2013

- Project: single view (Telecom billing product)
- Wrote low-level designs for new code packs and updated designs for existing code packs.
- Made code packs inter-operable in different versions of single view without impacting the functionality.
- Designed Telecom BSS solutions.
-

IBM-Product Support Engineer

DEC2011-JUL2013

- Worked in a Balance Transfer Application for Vodafone Spain.
- Developed tools for monitoring and health Checkups which grouped log data for server health reports
- Fixed common operational bug related to node failure.

PIVOTAL CONSULTANCY-System Engineer

JAN2011-DEC2011

- Built plug-and-play modules for WebSphere commerce servers solving purposes like Billing, Add-To-Cart, Shipping.
- Worked on providing self-developed solutions for custom WebSphere servers.
- Preparation of sales proposal and RFP adjustment.

IBM-Software Developer

AUG2008-JAN2011

- Worked as a Back-End Developer for LAPU TopUp recharge application.
- Built a Configuration Tool to interact with an intelligent network (IN).

EDUCATION

- Deep Learning Specialization by deeplearning.ai.
- Certified Scrum Master
- Bachelor of Engineering (**Computer Science**) – **MD university India**
- Currently pursuing a Master's in Computer Science with a specialization in Artificial Intelligence.