

Anshul Sharma

(+91) 9871771898 | anshuls235@gmail.com | [Kaggle](#) | [Medium](#) | [GitHub](#) | [LinkedIn](#)

Full-Stack Data Scientist with 9 years of experience in designing, building, and maintaining AI applications, including Generative AI, that enhance customer experiences, optimize business processes, and provide insights that drive innovation business growth. Proficient in building data pipelines, predictive modeling and seamless deployment of ML architecture using Cloud MLOps.

WORK EXPERIENCE

Senior Data Scientist

JUL 2022 — Present

Gartner

Gurgaon, Haryana

- Focused on driving seller productivity by applying **Data Science, ML, GenAI, Advanced Analytics**, and **BI** to real-world challenges. This includes ranking enterprises by predicted growth, identifying whitespace, quantifying opportunities, and analyzing seller workload to empower more strategic, data-driven decisions.
- Led a series of **GenAI** projects to tackle real business challenges, from building the Recruitment Studio—a powerful resume screening and scoring tool with **LLMs (Mixtral 8x7B, Llama2 70B)**—to fine-tuning the **Mistral 7B Instruct** model with a **RAG** setup for a company-specific **QA chatbot**. These initiatives cut down manual HR review time, boosted internal knowledge access by **40%**, and slashed response time for complex queries by **60%**, bringing measurable impact across both HR and internal support.
- Automated an ML workflow for quarterly growth forecasts, cutting turnaround time from **14 days** to **4 hours** and conserving over **10,000 person-hours** monthly.
- Orchestrated a Pyspark-based optimization of data pipelines, shrinking data processing time from **12 hours** to **12 minutes**, and instituted a robust MLOps framework via MLFlow to mitigate data and concept drift.
- Engineered a neural network-based workload index for **5,000 sellers**, revolutionizing territory investment strategies and decision-making processes.
- Automated and streamlined the NPV model for territory investment, eliminating **20 Excel sheets** and **3 weeks of manual labor** monthly, reducing the computation time to a mere **5 minutes**.

Senior Data Scientist

JAN 2021 — JUL 2022

S&P Global

Gurgaon, Haryana

- Pioneered the development of a Vehicle Miles Traveled product, generating **\$300k in revenue** during phase-1 and extending support to **20 global automotive markets** through LightGBM/Random Forest algorithms.
- Architected a **CI/CD-integrated, serverless workflow** leveraging AWS services, serving as a blueprint for other teams transitioning to serverless architectures.
- Led the Data Acquisition Team in scraping **50+ websites** and deploying large-scale web crawlers, amassing **gigabytes** of valuable data.
- Orchestrated the acquisition and transformation of data from **10+ disparate sources**, including Oracle, Hadoop, and MongoDB, employing advanced analytics to render the data actionable for stakeholders.

Developer

AUG 2015 — DEC 2020

SAP Labs India

Gurgaon, Haryana & Bengaluru, Karnataka

- Engineered **ETL pipelines** to automate data transfer from source files to **SAP BW**, and implemented **REST APIs**, doubling performance metrics and enhancing reporting capabilities.
- Integrated a **Random Forest Classifier** to replace a manual rule-based mapping system, achieving a **200% speed increase** in deriving automotive attributes. Additionally, contributed to **customer demos** and **product design discussions**, providing data-driven insights for product development.

TECHNICAL SKILLS & EXPERTISE

ML Frameworks & Data Science	: scikit-learn, XGBoost, LGBM, PyTorch, Keras, Hugging Face, Langchain, LlamaIndex, Pandas, NumPy
Generative AI	: LLMs, LLM Fine-tuning, RAG, Prompt Engineering
Programming Languages & Tools	: Python, SQL, R, PySpark, Spark, Git, Shell scripting, VS Code, PyCharm, Jupyter, Anaconda, KNIME
Cloud & MLOps	: AWS (SageMaker, Lambda, Step Functions, ECS, ECR, Batch, API Gateway, CloudWatch), Azure (App Service, ADF, Cognitive Services), Databricks, Docker, Terraform, Jenkins, MLflow, GitLab CI
Data Visualization	: Power BI, Tableau, Plotly, Matplotlib, Seaborn
Web Development	: Flask, Django REST framework, Streamlit, Gradio, HTML, CSS, Bootstrap, Dash

EDUCATION

BTech in Electronics & Communication, Delhi Technological University

2011 — 2015

Machine Learning Nanodegree, Udacity

2019

Data Analyst Nanodegree, Udacity

2018