

Data Scientist specializing in AI, MLOps and LLM’s with 5+ years of experience designing scalable production grade ML systems and building data to decision pipelines that deliver tangible business outcomes.

SKILLS

Data Science & Machine Learning: Supervised-Unsupervised-Deep Learning, Hypothesis Testing, AB Testing, LLM, AI Agents, CV, MLOps, MCP

Technical Toolkit: MYSQL, Python, Pandas, NumPy, Pytorch, Tensorflow, Docker, Terraform, CDK, Github, AWS(S3, Lambda, IAM, EC2, SageMaker, ECS, RDS, DynamoDB, Glue, Athena, QuickSight, Rekognition)

Data Engineering & Analytics: DataLake, Data Visualisation, ETL Pipelines, Tableau

EXPERIENCE

Data Scientist ← Associate Data Scientist ← Associate Consultant

Aug '22 - Present

Minify - Premier Partner AWS in Data & AI

- To streamline and scale proposal generation for an **Australian client**, built a **containerized RAG system** using **Docker**, **Amazon ECS Fargate**, and **Bedrock**. The solution automated document ingestion and retrieval, reducing manual effort by **40%**.
- To ensure accurate crop identification for subsidy allocation by the **MP Government**, developed a **scalable inference pipeline** using **AWS Lambda**, **SQS**, and **Step Functions** for a **Rekognition Custom Labels model**. Processed **33 million crop images** with **75% accuracy**, enabling data-driven fund distribution to farmers.
- To automate enterprise knowledge retrieval for a **US client**, led the development of a **cloud-based RAG solution** using **AWS CDK**, **Lambda**, **Step Functions**, and **Bedrock**. Optimized data indexing with **OpenSearch**, improving processing speed by **30%** and enhancing chatbot responsiveness.
- To centralize health data and automate reporting for a **Healthcare organization**, designed an **ETL pipeline** using **AWS Glue** and **Athena**, consolidating **600+ MongoDB collections** into a **Data Lake** on **S3**. Built **35 QuickSight dashboards** for **NABH hospitals**, improving high-level decision-making.
- To improve production planning for an **FMCG client**, developed a **Time Series Forecasting solution** using the **XG Boost algorithm** on **SageMaker**, predicting sales for **100+ SKUs** over **90 days** with **91% accuracy**. Implemented **end-to-end MLOps pipeline** for automated training and deployment, enhancing production efficiency by **35%**.

Data Science Consultant | HRS Groups

Apr '22 - Jun' 22

- To identify bottlenecks in global payment operations of hotels, analyzed **invoice collection** and **virtual credit card acceptance ratios** for over **100K hotels** using **AWS Athena**. Visualized results in **MicroStrategy** and created **performance dashboards**, reducing manual reporting workload by **~20% weekly**.
- To analyse key factors influencing payment adoption, built an **XGBoost model** with **62% accuracy** to predict major drivers impacting **virtual credit card acceptance** across hotel partners.
- To understand payment behavior patterns, applied **K-Means clustering** to segment hotels based on **chain affiliation, contract type, and amenities**, generating insights for improving acceptance rates.

Operations Associate | Aaryahaan International Pvt Ltd

Nov'17 - Mar' 21

- To unify consumer footfall analytics, managed and executed the **end-to-end project**, covering **6 large scale Retail Hubs**.
- To enable centralized performance visibility, collaborated with a **5-member cross-functional team** of **Data Scientists and Product Managers** to develop **real-time** consumer behavior dashboards on **Tableau**, resulting in a **10-15% MoM increase** in repeat consumer base.
- To predict the conversion rate from **passerby to shopper**, built **Customer Conversion Random Forest** models with an average accuracy of **86%**.

PROJECTS

Customer Purchase Analytics

- To help businesses identify potential customers and improve marketing efficiency, developed a **Random Forest** classification model that predicts purchase likelihood based on customer **demographics** such as **age, gender**, and **credit score**, enabling targeted and improved lead prioritization.

Unseenergy

- Worked on the business idea of collecting **energy consumption data** to visualize and optimize its usage. Generated personalized recommendations using the **SARIMA** model to give insights on how energy can be saved.

EDUCATION

PG Tech Program in AI-ML	Plaksha University - UC Berkeley - Purdue	Jul' 2022
Bachelor in Computer Applications	Amity University	Jun' 2017

CERTIFICATIONS

- AWS Machine Learning Specialty - Built and deployed **MLOps solutions** using **SageMaker** and **serverless ML services**
- AWS Data & Analytics Specialty - Designed **data pipelines and analytics workflows** using **Glue, Athena, and Redshift**
- AWS Solutions Architect Associate - Architected **secure and cost-optimized AWS cloud solutions**