

Shashank Mathur - Data Science

Gurugram-India | +91-9928131452 | mathurshashank2299@gmail.com | [shashank-mathur](#)

SUMMARY

As an experienced **Data Scientist** specializing in **Machine Learning (ML)** and **Generative AI (GenAI)** technologies, I aim to leverage my expertise to drive actionable insights and **data-driven decisions**. With over two years of hands-on practice, including experience with **LLM-powered applications**, I excel in solving complex problems and collaborating on transformative data projects. My passion lies in pushing the boundaries of AI applications, enhancing **predictive analytics**, and **developing innovative data processing modules**.

EDUCATION

JIET Jodhpur

B-Tech in Computer Science

Jodhpur

Graduation Date: Jan 2022

SKILLS & INTERESTS

Programming Skills: Python, HTML, CSS, Java.

Database: SQL, PostgreSQL, SQLite, Tiny DB.

Technologies/Frameworks: Django, Flask, Pandas, Open AI, Sklearn, Apache-Pyspark, Spring Boot, Tensorflow.

Version-Control/Cloud: Git & Github, Azure, AWS, Databricks

Domain-Knowledge : Backend-Development, Data Science, Machine Learning, Large Language Models (LLMs), Extract Load Transform (ETL), Generative AI, Prompt Engineering

WORK EXPERIENCE

Capgemini Technology Services

Senior Associate Software Engineer

Gurugram India

Jan 2022 - Finish Date

- Collaborated with clients from **Fashion Retail, BFSI, and Non-profit sectors**, conducting thorough consultations to grasp their unique requirements and deliver tailored **machine learning and AI solutions**.
- Engineered an **end-to-end ETL pipeline** for taxi services on **Azure Databricks**, utilizing Apache Spark and **Azure Delta Lake (ADL)** for efficient data processing and reporting.
- Applied **ensemble techniques** proficiently to address intricate **classification and regression challenges**, resulting in the creation of highly accurate **predictive models**.
- Successfully achieved a notable 40% reduction in marketing costs by implementing lead-scoring methodologies.
- Employed Language Models (LLMs)** in chatbot development and harnessed **Generative AI (Gen AI)** across various applications to elevate user experience and interaction quality

Bhaskaracharya Institute for Space Applications and Research

Machine Learning intern

Remote

Jan 2021 - Jan 2021

- Utilized cutting-edge algorithms to enhance accuracy and efficiency of image processing tasks, resulting in a 20% reduction in processing time for **Machine Learning and Computer Vision operations**.
- Implemented a new data preprocessing technique that improved model performance by 15% when training **custom object detection models**.
- Collaborated with cross-functional teams to integrate advanced **Image Processing techniques** into existing workflows, increasing overall productivity by 30%.

YellowFame

Machine Learning and Backend Developer

Jodhpur Rajasthan

Jan 2021 - Jan 2024

- Developed and implemented **machine learning algorithms** to enhance the accuracy of virtual jewellery try-ons by 30%, resulting in a 20% increase in user engagement.
- Utilized **Django and PostgreSQL** to optimize database performance, reducing API response time by 40% and increasing app efficiency.

PROJECT EXPERIENCE

ETL Pipeline for Taxi Services

- Built and optimized an **end-to-end ETL pipeline** using Apache Spark on **Azure Databricks** for taxi service data, resulting in a 40% increase in **data processing** speed and accuracy.
- Extracted, preprocessed, and analyzed data from **Azure Delta Lake** Blob Storage using Databricks, leading to the identification of key trends that drove a 15% increase in operational efficiency.
- **Automated daily job execution** using Databricks Jobs and **Azure Data Factory**, reducing manual workload by 70% and ensuring consistent data delivery.

Q&A Bot – Networking

- Engineered an **Azure OpenAI**-powered solution, **RAG-based architecture**, which is a combination of a Summarizer and Q&A system for insightful file analysis.
- Developed a PDF parsing solution for diverse file formats using **Azure Doc Intelligence**, created embeddings using the **OpenAI ADA** model and stored in Redis DB for efficient search.
- Established **CI/CD pipeline** for Automation using **Azure DevOps** for seamless model deployment, and used Azure Cosmos DB for session management.
- Utilized Azure Cosmos DB for comprehensive log management and implemented load balancing for more than 5K users with **Azure API Management** for error mitigation.
- Our solution helped the customer to improve their productivity by approximately 15%.

Fashion Retail Demand Forecasting

- Developed a **demand forecasting** system for the **fashion retail industry** to optimize inventory management.
- Performed detailed analysis of historical sales data, identifying patterns, and crafting exogenous factors to enhance the model's performance and trend capture.
- Utilized **Prophet** and **SARIMAX** models and applied **hyperparameter tuning** to enhance model performance.
- Automated the pipeline using **Databricks workflows** and deployed the model on **MLFlow**, providing a user-friendly endpoint.

Donor Classification

- Developed a classification model to optimize marketing expenditure and donor acquisition for **250+ nonprofit organizations**.
- Performed data preprocessing on extensive datasets stored on **AWS S3** and employed **ensemble models** on Databricks for donor classification.
- Managed model deployment with **MLflow** and leveraged the seamless execution and monitoring of pipelines through **Databricks workflows**.
- Enhanced donor acquisition strategies, reduced costs, and reported progress directly to the **Vice President through JIRA**.

CERTIFICATIONS

- **Applied Data Science Pyspark** | Coursera-DataBricks | Jan 2023
- **Data Science fundamentals pyspark** | Coursera-DataBricks | Jan 2023
- **Apache Spark (TM) SQL for Data Analytics** | Coursera-DataBricks | Dec 2022
- **Machine Learning (Star Performer)** | Cipher School | May 2020 to June 2020