

Profile Summary

- Data Engineer with hands-on experience in building data pipelines using Azure Data Factory, Databricks, Apache Spark (PySpark), and SQL. Skilled in ETL/ELT workflows, data modeling, and analytics engineering using dbt. Experienced in Python-based data processing, ML model development, and integrating AI solutions using LLMs and RAG pipelines. Currently interning at KPI Partners with practical exposure to enterprise-scale data engineering projects.
- Experienced in building end-to-end academic projects, including a **Flask-based Library Management System**, a **Business Data Management project** for Write Wing Advertising, a **machine learning model** for predicting the success of bank telemarketing, and **AI-powered enhancements** to the IIT Madras learning portal. Skilled in **data manipulation, model development**, and developing functional prototypes.

EDUCATION

- **BS in Data Science and Applications**
With a minor in Computer Systems
Indian Institute of Technology
Madras, Chennai, Tamil Nadu

IIT MADRAS

Business Data Management Project - Mentor

Some of the responsibilities and activities:

- Guided 100+ students through the complete Business Data Management (BDM) project lifecycle, including data collection, cleaning, modeling, and report preparation.
- Provided structured mentorship on SQL, data analysis, Excel workflows, and database design concepts.
- Helped learners interpret business requirements and translate them into data-driven insights and solutions.
- Ensured adherence to academic guidelines while encouraging best practices in documentation and presentation.

Projects

1. Library Management System (HTML/CSS, Python, Flask, SQL)

Developed a full-stack Library Management System enabling efficient book tracking, user

- **12th Grade 95.00%**
South Point High School, Kolkata, West Bengal

TECHNICAL SKILLS

- **Languages:** Python, MySQL, C, C++, JavaScript, HTML/CSS
- **Data Engineering Tools:** Azure Data Factory (ADF), Azure Databricks, Apache Spark (PySpark), dbt, SQLAlchemy, Psycopg2
- **Libraries:** Scikit-learn, PyTorch, NumPy, Pandas, Matplotlib, Seaborn
- **LLMs, AI Agents & RAG:** RAG Pipelines, LLM Integration (Gemma, Ollama models), Embedding Models, Vector Databases (FAISS / ChromaDB), Document Chunking, Query Embedding & Similarity Search, Building Simple Task-Oriented AI Agents
- **Developer Tools:** Git, GitHub, VS Code, PyCharm, Jupyter, Postman

management, and automated notifications. Built secure role-based access for admins and users, designed a complete book and user management workflow with real-time updates, and implemented a smooth borrow–return system with automated reminders, improving overall usability and circulation efficiency

2. Business Data Management Project(Python, SQL, Numpy, Pandas, Matplotlib)

Executed an in-depth analysis of customer, marketing, and website performance data for WriteWing Advertising, uncovering revenue patterns, service utilisation trends, and satisfaction gaps. Performed full data preprocessing, segmentation, and marketing effectiveness analysis, delivering insights and recommendations that supported improved customer acquisition, better service engagement, and more informed business decision-making.

3. Machine Learning Project – Bank Telemarketing Success Prediction (Numpy, Pandas, Scikit-Learn, xGBoost)

Built a predictive model to determine customer subscription likelihood, achieving a top 12% rank on the Kaggle leaderboard. Conducted extensive data preprocessing, feature engineering, exploratory analysis, and model evaluation across multiple classifiers, producing an optimized

CERTIFICATIONS

- **AWS Academy Graduate – Cloud Foundations**
 - Gained a strong understanding of cloud computing fundamentals and AWS global infrastructure.
 - Learned core AWS services such as EC2, S3, RDS, and IAM, along with security and compliance essentials.
- **Databricks Certified Data Engineer Associate**
 - Earned Databricks Data Engineer Associate certification, demonstrating proficiency in Spark, ETL pipelines, data modeling, and production-grade data engineering workflows practices.
- **SQL (Basic) – HackerRank**
- **Introduction to Programming Using Python – HackerRank**
- **Data Mining – Nptel**
 - Acquired knowledge of data mining techniques such as

workflow that delivered strong predictive performance and actionable insights into key factors influencing customer conversion.

4. AI-Enhanced Learning Portal – Software Engineering Project (LLM, RAG, Embeddings)

Designed and integrated AI-powered features into the IIT Madras learning portal to transform it into an intelligent study assistant. Added capabilities such as automated lecture summaries, doubt resolution, assignment evaluation, and context-aware responses through LLM-based workflows. Implemented retrieval-augmented generation, structured prompt flows, and AI endpoints, resulting in improved student engagement and faster doubt resolution.

5. Building Data Pipeline using Azure Medallion Architecture

Built a data pipeline to analyze large sensor data collected from water bodies across different European countries over several years using Azure Services and SQL Server to generate visualizations.

The pipeline is designed to ingest sensor data, process it, and move it to the destination, utilising various Azure services such as Azure SQL Database, Azure Logic Apps, Azure Storage Accounts, Azure Data Factory, and Azure Databricks. The final transformed data is visualized using Power BI, the most demanding Business Intelligence tool. The pipeline is structured to be modular and customizable, which

- clustering, classification, and association rule learning.
- Learned how to extract meaningful patterns and insights from large datasets.

allows additional Azure services to be seamlessly integrated to expand its functionality