

# Somil Doshi

713-572-5426 | somildoshi1202@gmail.com | [linkedin.com/in/somil12doshi/](https://linkedin.com/in/somil12doshi/) | [github.com/somildoshi12](https://github.com/somildoshi12)

## Education

### University of Houston

*Master of Science in Engineering Data Science*

**Aug 2024 – May 2026**

GPA: 3.83/4.0

### University of Mumbai

*Bachelor of Engineering in Information Technology*

**Aug 2020 – May 2024**

GPA: 3.84/4.0

## Professional Experience

### Cullen College of Engineering | University of Houston

**Mar 2025 – Present**

*Research Assistant*

- Structured a multi-agent systems using **LangChain** and **LangGraph** to design reasoning pipelines, context-aware planning flows, and autonomous tool integration, improving orchestration efficiency by 35% across multiple projects.
- Implemented retrieval-augmented generation pipelines with **Databricks** and Pinecone, fine-tuned LLMs on the domain specific datasets, and optimized embeddings for semantic search, improving query relevance scores by 22%.
- Deployed services with **PostgreSQL**, **Docker**, **Kubernetes**, enabling scalable and secure model access for research.

### The Sparks Foundation

**Sept 2022 – May 2023**

*Data Analyst Intern*

- Consolidated 50K+ energy usage records in **MongoDB**, **Databricks** with Airflow orchestration, applying the DBT transformations to align, improve forecasting reliability by 19% across multiple downstream analytics workflows.
- Configured **LSTM** forecasting pipelines in Databricks and **AWS SageMaker**, integrated with **Kafka** and validated through Airflow DAGs, sustaining 92% accuracy in backtesting and reducing prediction error margins by 18%.
- Designed **Power BI** dashboards using DAX queries to visualize seasonal demand shifts and generate anomaly alerts, enhancing decision-making capabilities for the analytics team.

### TCR Innovations

**Jan 2022 – Aug 2022**

*Data Engineer Intern*

- Introduced ETL pipelines in Python with **AWS RDS** integration, managing 80K+ drilling records and enforcing schema rules, reducing refresh cycles by 34% and improving reliability of downstream workloads.
- Deployed predictive APIs powered by **Random Forest** and **XGBoost** models, linked to **React.js** interfaces, sustaining 89% accuracy while reducing manual intervention for operational teams.
- Automated cloud deployments with Docker, **AWS Lambda**, ensuring stable builds across staging and production.

### Multicon IOTW

**Aug 2021 – Dec 2021**

*Student Data Engineer*

- Implemented Python and **PostgreSQL** ETL workflows to clean 50K+ student performance and attendance records, increasing data reliability for departmental use by 29%.
- Generated SQL-based reports and **Excel** dashboards, enabling faculty to track semester-wise performance trends and improving academic progress monitoring efficiency by 34%.
- Streamlined deployment pipelines with **Docker**, automating ETL workflow execution and ensuring consistency for departmental data projects.

## Academic Projects

### Predictive Maintenance for Oil Wells | GRU Neural Networks, FastAPI, React.js, PostgreSQL, Docker, Streamlit

- Forecasted equipment failures 7 days earlier with GRU models, sustaining 91% accuracy on drilling sensor data.
- Reduced unplanned downtime by 28% and optimized maintenance scheduling across 5,000+ simulated well operations.

### Brain Tumor MRI Classification | 3D CNNs, PyTorch, Flask APIs, MongoDB Atlas, Docker

- Trained 3D CNN models on MRI volumes, achieving 94% accuracy in classifying multiple brain tumor subtypes.
- Exposed predictions through Flask APIs with MongoDB Atlas logging, reducing diagnostic reporting time by 30%.

### Construction Safety Monitoring | YOLOv5, PyTorch, Azure Cosmos DB, Docker, Flask APIs, PowerBI (DAX)

- Designed YOLOv5 models for helmet and vest detection, reaching 96% accuracy on site surveillance datasets.
- Logged and visualized incidents and reports with Streamlit, improving reporting efficiency by 33% in simulations.

## Technical Skills

- Programming Languages & Tools** - Python, R, Java, SQL, JavaScript, React.js, Git, MS Excel
- Frameworks & Platforms** - PyTorch, TensorFlow, YOLOv5, LangChain, FastAPI, Flask, AWS SageMaker
- Data Engineering & Deployment** - Airflow, DBT, PySpark, Docker, Kubernetes, GitHub Actions, ETL Pipelines, Jenkins
- Databases & Visualization** -PostgreSQL, AWS RDS, MongoDB Atlas, Azure Cosmos DB, MySQL, Power BI, Tableau
- Relevant Skills** - Analytical Thinking, Communication, Team Collaboration, Project Management, Continuous Learning