

Sachin Ramesh Shet

✉ shet.sa@northeastern.edu | 🌐 SachinShet73 | 📧 sachin-shet | 📞 +617-505-8176

Education

Northeastern University

Master of Science in Computer Software Engineering (GPA: 3.667/4.0)

Boston, MA

September 2024 – Present

- Advanced Coursework: Data Science Engineering Methods and Tools, Program Structures and Algorithms, Database Systems

Experience

Infosys

Senior Systems Engineer

Promoted from Systems Engineer (November 2021 – September 2023)

Bangalore, India

October 2023 – February 2024

- Architected end-to-end **Extract Transform Load (ETL) pipeline** using **Azure Data Factory**, implementing lookup activities for batch validation and orchestrating automated data flow between server, **Linux** environment, and **Azure Blob Storage** for **2GB+** daily data volume
- Led technical discussions with international clients to transform complex business requirements into scalable solutions, resulting in successful submission of 3 major deliverables ahead of schedule
- Developed dynamic **Python** automation scripts processing **55,000+** rows daily, reducing processing time by **60%** through automated table creation and data insertion
- Engineered resilient data migration workflows between **Azure Blob Storage** and **Snowflake**, maintaining **99.9%** data accuracy across **180+** tables with automated validation
- Orchestrated complex ETL workflows using **Apache Airflow**, reducing manual intervention by **90%** and achieving zero data loss across 6-month deployment
- Received prestigious **Blue Badge** award (top 5% of engineers) and earned **two promotions** within 2.5 years for my technical contributions

RoveLabs

Machine Learning Intern

Davangere, India

April 2020 – September 2020

- Authored a real-time candidate verification system using **OpenCV** and **dlib**, achieving **95%** accuracy in face detection and maintaining **30 FPS** processing speed
- Implemented concurrent processing for multimodal data analysis, integrating video streams and audio pattern detection with **TensorFlow** and **librosa**, resulting in less than **2-second** latency
- Engineered an audio anomaly detection system achieving **90%** accuracy, while reducing false positives in suspicious activity detection by **40%**

Projects

Real-time Social Media Sentiment Analysis Pipeline | *Kafka, Spark, Docker, PostgreSQL, Elasticsearch*

- Engineered a real-time data pipeline using **Twitter API** and **Apache Kafka**, processing **1,000+** posts/hour, reducing brand monitoring time from days to minutes and enabling immediate responses to customer sentiment
- Developed real-time visualization dashboards using **Kibana**, allowing teams to track sentiment trends and launch targeted marketing initiatives, driving data-driven decision-making
- Implemented distributed data processing using **Apache Spark**, achieving **70%** sentiment analysis accuracy with **2-second** processing latency
- Designed a hybrid storage solution combining **PostgreSQL** for structured data and **Elasticsearch** for efficient text search, reducing query response times to under **3 seconds**
- Containerized the entire application stack with **Docker**, ensuring consistent deployment across development environments and achieving a **90%** deployment success rate

Technical Skills

Programming Languages & Core: Python (NumPy, Pandas), Java, SQL

Data Engineering: Apache Kafka, Apache Spark, Apache Airflow

Cloud & Infrastructure: Azure Data Factory, Azure Blob Storage, Docker, Linux, CI/CD

Data Storage & Analytics: Snowflake, PostgreSQL, Elasticsearch, MongoDB

Data Visualization: Kibana

Machine Learning: TensorFlow, OpenCV, dlib, librosa