

Shardul Chavan

+18573135138 | shardulc84@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

EDUCATION

Northeastern University - Master of Science, Information Systems, **GPA: 3.76/4.00**

Sep 2022 — Dec 2024

University of Mumbai - Bachelor of Engineering, Computer Science and Engineering, **GPA: 3.80/4.00**

Aug 2019 — May 2022

WORK EXPERIENCE

Crewasis, Remote - Data Science – Engineer Intern

Jan 2025 — Jun 2025

- Spearheaded **ML pipeline processing 2M+ text records** from social/news feeds, implementing sentiment analysis and topic classification to improve marketing insights accuracy by 15%
- **Architected automated model retraining** system using AWS Lambda and EC2, reducing manual ML operations workload by 70% through scheduled jobs and threshold-based triggers for model drift detection
- **Integrated AWS Bedrock and OpenAI APIs** to enhance NLP outputs with contextual insights, creating Python-based data aggregation service that reduced report generation time from 2 hours to 15 minutes
- **Delivered Tableau dashboards** visualizing NLP-driven sentiment trends and topic clusters, empowering marketing teams to track campaign performance metrics in real-time, reducing ad-hoc reporting requests by 35%
- **Led CI/CD implementation using GitHub Actions and Terraform** in Agile environment, establishing automated testing and deployment workflows, reducing production bugs by 20% and improved team deployment velocity

Skyworks Solutions Inc., Boston, MA - Data Engineer Co-op

Jan 2024 — Jun 2024

- **Designed 25 Apache Airflow pipelines** to automate RF semiconductor test data ingestion from multiple vendors, reducing data availability latency from 6 hours to under 1 hour for downstream analytics teams
- **Created Python-based parsing tool** to extract and transform vendor BOM data from PDF files into structured formats, saving test engineers 5 hours daily and standardizing data inputs across 8 different vendor formats
- **Instituted Great Expectations for data validation** across ETL pipelines, defining 50+ critical data tests that reduced data quality incidents by 80%
- Streamlined SQL queries and **executed incremental data loading strategies** in Azure Synapse, improving Power BI dashboard performance by 60% for 3 business units tracking operational KPIs
- Containerized data processing services using Docker and **implemented monitoring with Prometheus**, achieving 99.5% uptime through enhanced alerting and retry mechanisms

Northeastern University, Boston, MA – Graduate Teaching Assistant and AI Research Engineer

Sep 2023 — Dec 2024

- Engineered SnowGPT, a natural language to **SQL translation tool using LangChain** and Snowflake, supporting non-technical users to query databases through conversational interface with 85% accuracy on common queries
- **Orchestrated PySpark data processing pipeline** for insurance risk modeling dataset of 1.2 million records, implementing feature engineering techniques, improving model prediction accuracy from 72% to 92%
- Deployed machine learning models as containerized microservices using Docker and FastAPI, creating REST endpoints for real-time inference and implementing model versioning for safe rollbacks
- **Mentored 50+ graduate students in AI/ML projects** through hands-on guidance, code reviews, and deployment best practices, enabling 10 successful capstone completions

Accion Labs, Mumbai, India - Software Engineer Intern

Jan 2022 — Jul 2022

- Integrated REST API connection between Large Language Models and ServiceNow platform, enhancing chatbot natural language understanding capabilities and improving ticket routing accuracy by 40%

PROJECTS

Real-Time News Aggregation Platform | Python, FastAPI, Docker, Google Cloud Platform, Pinecone, PostgreSQL

- Constructed news aggregation system processing articles from 10 RSS feeds using vector embeddings for content similarity, delivering personalized updates every 8 minutes with 90% user relevance score
- Deployed containerized microservices on Google Cloud Run with auto-scaling, supporting 100 concurrent users maintaining sub-second response times

AWS-Based YouTube Data Analytics | IAM, AWS Lambda, AWS Glue, Amazon Redshift, Apache Spark, QuickSight

- Architected serverless ETL pipeline using AWS Lambda for parallel API extraction and Glue Data Catalog for schema discovery, processing 80K+ daily video interactions through S3 staging
- Implemented incremental data loading and partitioning strategies in Redshift, reducing query execution time by 75% and building QuickSight visualizations enabling to identify content trends across demographics

Retail Sales Data Pipeline | Azure Data Factory, Apache Spark, Power BI, Python, SQL

- Built end-to-end analytics pipeline processing 25 million historical sales transactions using Azure Data Factory and Databricks, designing star schema dimensional model with fact tables for sales and inventory tracking across 50 store locations
- Leveraged the processed data to develop ensemble forecasting system combining ARIMA, and XGBoost models, engineering features from historical sales patterns, improving demand prediction accuracy by 30% and reducing overstock incidents by 40%

SKILLS

Programming Languages	: Python, Java, SQL, R, C++, JavaScript, Scala, Bash, SAS
Data Engineering	: Apache Spark, Apache Airflow, dbt, ETL/ELT, Databricks, PySpark, Data Modeling, Kafka
Databases	: Snowflake, PostgreSQL, MongoDB, MySQL, Redshift, DynamoDB, BigQuery, Redis
Cloud Platforms	: AWS (EC2, Lambda, S3, Redshift, Glue, RDS, CloudWatch), Azure, Google Cloud Platform
Business Intelligence	: Tableau, Power BI, QuickSight, Excel, Looker, R Shiny
Machine Learning	: MLOps, Feature Engineering, Model Deployment, NLP, Deep Learning, Computer Vision, Hugging Face
Web Frameworks & Libraries	: PyTorch, TensorFlow, Keras, Scikit-learn, Matplotlib, Seaborn, Pandas, NumPy, NLTK, FastAPI, Flask, React
DevOps & Infrastructure	: Docker, Kubernetes, Terraform, Jenkins, GitHub Actions, CI/CD, Linux
Data Governance	: dbt Tests, Data Quality, Great Expectations, Metadata Management, Compliance, Security Best Practices
Methodologies	: Agile/Scrum, JIRA, Confluence, Cross-functional Collaboration