

SANJANA PATNAM

(857)605-1438 | patnam.sa@northeastern.edu | Boston, MA | [LinkedIn](#) | [Github](#) | [Portfolio](#)

SUMMARY

Data Analytics Engineering grad student skilled in cloud-native data pipelines, Python, SQL, AWS, and Databricks. Passionate about scalable analytics, team collaboration, and solving real-world data challenges.

EDUCATION

Northeastern University, Boston, MA

May 2027

Master of Science in Data Analytics Engineering; GPA: 3.8/4

Relevant Coursework: Data Management, Data Computations and Visualization, Data Mining, and Machine Learning

Geethanjali College of Engineering and Technology, Hyderabad, India

May 2024

Bachelor of Technology in Computer Science and Engineering

Relevant Coursework: Data Structures, AI & ML, DBMS, Big Data Analytics, Cloud Computing, and Web Technologies

SKILLS

Languages: Python, SQL, R

DevOps & Infra: Docker, Kubernetes, Terraform, Git

Cloud Platforms: AWS (S3, Redshift, Glue), GCP, Azure

Testing: Pytest, Integration Testing, CI/CD validation

Big Data: Apache Spark, Hadoop, Databricks

Monitoring: CloudWatch, Airflow UI, custom alerting

Orchestration: Airflow, Prefect, Jenkins

Streaming: Apache Kafka

Databases: MySQL, PostgreSQL, MongoDB

Certifications: AWS DevOps (Advanced), HackerRank SQL

Soft Skills: Collaboration, Attention to detail, Communication

WORK EXPERIENCE

Quadrant resources Pvt.Ltd., Hyderabad, India

May 2024 – Nov 2024

Web Data Analyst

- Managed **ETL pipelines** for NLP and search data to support **AI model** training workflows
- Annotated and processed over **5,000 records weekly**, using **transformation techniques** that increased **data precision** by up to 20%
- Implemented basic **monitoring logic and manual alerting** scripts to flag **data anomalies** during ETL runs
- Collaborated with ML engineers and QA teams to maintain over **98% pipeline accuracy**

1Stop, Hyderabad, India

July 2022 – Sep 2022

Student Intern

- Designed and optimized **MySQL schemas and queries**, boosting **query performance by 20%**
- Built and deployed an **ETL-enabled e-commerce platform** with real-time **inventory and user management**
- Integrated **XAMPP stack** for deployment and debugging, reducing **platform latency by 25%**
- Implemented basic **CI/CD pipeline** to automate testing and push fixes cycles in a small-team Agile environment.

TECHNICAL PROJECTS

Precision Oncology Data Integration and Analytics System, Northeastern University

April 2025

- Built **SQL and NoSQL databases** to integrate clinical, genomic, and imaging datasets
- Developed **data pipelines** using Python and **Databricks** on **AWS** for unifying heterogeneous data
- Designed **Power BI dashboards** for real-time KPIs on treatment outcomes, enabling clinicians to make data-driven decisions faster and **improving response time** to patient changes

Automated University and Research Data Crawler

March 2025

- Built a **Python-based web scraper** and deduplication pipeline for ranking university data using **ETL workflows** and citation-based metrics.

Patient Clustering for Medical Insights, Northeastern University

Feb 2025

- Built **preprocessing pipelines** and derived health indicators for clustering workflows, improving **segmentation by 18%**
- Built **data visualization dashboards** to validate clusters and assist with **medical interpretation**

Fake News Detection, Geethanjali College of Engineering and Technology

March 2024

- Designed a scalable **ML data pipeline** using **TF-IDF features** and **Logistic Regression** for article classification
- Extended pipeline with **ensemble models**, achieving **90%+ accuracy** and **0.85 F1 score** for binary classification

VOLUNTEER & EXTRACURRICULAR

- Member, **DAESO (Data Analytics Engineering Student Org)** at Northeastern University
- Attendee, **PyData Boston AI Agent & MCP Deep Dive Meetup (2025)**