SANJANA PATNAM

(857)605-1438 | patnam.sa@northeastern.edu | Boston, MA | LinkedIn | GitHub | Portfolio

EDUCATION

Northeastern University, Boston, MA **May 2027**
Master of Science in Data Analytics Engineering; GPA: 3.87/4
Relevant Coursework: Data Management, Data Computation and Visualization, Data Mining, MLOps, Applied NLP

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, Web Technologies

SKILLS

Languages: Python, SQL, R
Analytics & Visualization: Pandas, Tableau, Power BI, Matplotlib, Seaborn, Scikit-learn
ML & Analytics: Supervised & Unsupervised Learning, Regression, Classification, Clustering, Time Series Analysis, NLP
Databases & Big Data: MySQL, PostgreSQL, MongoDB, Apache Spark
Cloud & MLOps: Azure (ML, Storage, Compute, DevOps), AWS, GCP, Docker, Git, MLflow
Soft Skills: Data storytelling, Strategic thinking, Communication, Adaptability, Attention to detail
Exploring: Generative AI, LLMs, HuggingFace, Transformers
Certifications: AWS DevOps (Advanced Testing), HackerRank SQL (Advanced)

PROFESSIONAL EXPERIENCE

Quadrant resources Pvt.Ltd., Hyderabad, India **May 2024 – Nov 2024**
Web Data Analyst

- Processed and annotated over 5,000 records weekly to improve **training datasets** for NLP and search algorithms, boosting ad relevance by 20%
- Supported **machine learning pipelines** for query intent classification and contextual ad ranking, contributing to improved **model precision**
- Orchestrated **data validation and labeling techniques**, maintaining 98%+ accuracy and ensuring reliable datasets for downstream **ML model training**

TECHNICAL PROJECTS

End-to-End Real Estate Price Prediction System (Azure ML) **Nov 2025**

- Performed **EDA and feature engineering** on 20+ housing variables to examine key drivers influencing pricing behavior
- Trained and evaluated **Lasso, Random Forest, and XGBoost**, analyzing **accuracy–interpretability tradeoffs** using RMSE and R^2
- Applied **K-Means clustering** to identify regional pricing segments and distinct property behavior patterns
- Deployed batch inference workflows on **Azure ML** and built **Tableau dashboards** to analyze pricing trends and interpret model tradeoffs and outputs

Precision Oncology Data Integration and Analytics System, Northeastern University **April 2025**

- Built ETL pipelines using **SQL and NoSQL** to integrate clinical and genomic data for patient-level analysis
- Applied machine learning–based stratification to identify risk groups, improving outcome prediction accuracy by **22%**
- Designed **KPI dashboards** to support clinical insight generation and decision-making

Fake News Detection, Geethanjali College of Engineering and Technology **March 2024**

- Developed an **NLP classification pipeline** using TF-IDF and supervised learning to detect misinformation
- Evaluated model performance using **precision, recall, and F1-score**, analyzing error patterns across content categories
- Implemented reproducible preprocessing and training workflows to support consistent experimentation

ADDITIONAL EXPERIENCE

Northeastern University, Boston, MA **Jan 2026 – Present**
Instructional Assistant

- Diagnose recurring **system and workflow issues** in live environments, performing **root-cause analysis** to improve reliability and response consistency
- Track and resolve technical issues through structured escalation workflows, documenting outcomes to strengthen **system reliability and operational processes**