SANJANA PATNAM

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

SUMMARY

Analytical and results-oriented Data analytics engineering graduate student with hands-on experience in building end-to-end ML pipelines for classification, clustering, and predictive analytics. Skilled in Python, SQL, and statistical modeling with a strong focus on deploying scalable solutions. Passionate about applying AI to real-world problems and currently exploring LLMs and generative AI.

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

Programming & Data: Python, SQL, R
ML & AI: Regression, Classification, Clustering, Predictive Analytics, Time Series, NLP
Data Visualization: Pandas, Tableau, Power BI, Matplotlib, Seaborn, Scikit-learn
Databases & Big Data: MySQL, PostgreSQL, MongoDB, Apache Spark
Cloud & MLOps: Azure (ML, storage, Compute, DevOps, CDI & SDK), AWS (S3, SageMaker, Redshift), GCP BigQuery, Docker, Git, MLflow, Perfect
Certifications: AWS DevOps (Advanced Testing), HackerRank SQL (Advanced)
Exploring: Generative AI, LLMs, HuggingFace Transformers
Soft Skills: Data storytelling, Strategic thinking, Critical thinking, Communication, Adaptability, Attention to detail

WORK 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 with Azure Machine Learning **Nov 2025**

- Engineered **high-impact predictive features** and performed **in-depth EDA** on 20+ housing variables, identifying key drivers of property valuation
- Developed, tuned, and **compared ML models** (Lasso, Ridge, Random Forest, XGBoost), selecting the best model using RMSE- and R^2-driven evaluation pipelines
- Applied **Unsupervised learning** (K-Means) clustering to discover market segments and property behavior patterns, enhancing model interpretability and regional insights
- Deployed the final model as an **Azure ML batch endpoint** to enable scalable inference and **integrated predictions into Tableau for analytical storytelling**

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

- Built **relational and NoSQL databases** to unify oncology datasets and applied **predictive analytics** to assess treatment effectiveness
- Improved patient outcome prediction by **22%** through **ML-based stratification** and visual KPI dashboards
- Enabled **AI-driven patient stratification models**, enhancing outcome prediction for medical research

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

- Developed a supervised **ML pipeline** using TF-IDF features and Logistic Regression to classify fraudulent articles
- Improved accuracy using **ensemble models (Random Forest, Gradient Boosting)** with 90%+ accuracy and 0.85 F1 score
- Simulated fraud detection use case with **automated, scalable model evaluation functions**

VOLUNTEER & EXTRACURRICULAR

- Member, **DAESO (Data Analytics Engineering Student Org)** at Northeastern University