SANJANA PATNAM

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

SKILLS

Programming: Python, SQL
Data Analysis & Manipulation: Pandas, NumPy, Data Cleaning, Exploratory Data Analysis (EDA)
Machine Learning: Classification, Regression, Feature Engineering, Model Evaluation
NLP & GenAI: TF-IDF, LLMs (GPT-4, Claude, Gemini), RAG, Prompt Engineering
Data Visualization: Tableau, Power BI, Matplotlib, Seaborn, KPI Dashboards
Experimentation & Reproducibility: MLflow, DVC, Hyperparameter Tuning, A/B Testing Concepts
Data Engineering Exposure: ETL Pipelines, Apache Airflow, PostgreSQL, MongoDB, Apache Spark
Cloud & Deployment: GCP (Cloud Run, BigQuery), Docker, GitHub Actions, Azure ML, AWS

PROFESSIONAL EXPERIENCE

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

- Processed and validated **5,000+ records weekly** supporting NLP and search ML pipelines, improving dataset precision by ~20%
- Designed structured data cleaning and preprocessing workflows achieving **98%+ dataset accuracy**
- Prepared model-ready datasets and supported downstream **ML evaluation and optimization**
- Conducted exploratory analysis to identify data inconsistencies and improve training data quality

TECHNICAL PROJECTS

SavVio — AI Financial Advisory Platform (In Progress), Northeastern University

- Develop **LLM + RAG pipelines** using embeddings and pgvector retrieval for context-aware financial recommendations
- Engineer feature pipelines and deterministic scoring logic (budget, affordability, risk) to support **ML-based decision systems**
- Design **feature scoring logic and evaluation workflows** to assess recommendation quality and user response patterns
- Build and maintain **Airflow ETL workflows** to ingest and standardize financial, product, and review datasets
- Run **prompt optimization and model evaluation** experiments to improve response reliability and performance
- Track experiments using **MLflow and version datasets** with DVC to ensure reproducibility
- Deploy Dockerized FastAPI services via GitHub Actions CI/CD to GCP Cloud Run

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

- Integrated multi-source healthcare datasets using **SQL/NoSQL pipeline**s to create analysis-ready tables
- Engineered features enabling patient stratification and outcome **prediction analysis** that identified high-risk cohorts with improved interpretability for clinical decision support
- Validated data consistency and built reusable transformations for repeatable **modeling workflows**

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

- Built **NLP classification model** using TF-IDF and Logistic Regression, improving detection accuracy over baseline by 12%
- Evaluated **model performance** using precision, recall, and F1 metrics, achieving improved classification accuracy through iterative **feature tuning**
- Conducted error analysis to **refine preprocessing** and improve **classification accuracy**

ADDITIONAL EXPERIENCE

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

- Diagnosed **recurring workflow failures** through root-cause analysis and **structured debugging**
- Documented resolution patterns to improve **reliability of production-like analytics systems**

EDUCATION

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

Geethanjali College of Engineering and Technology, Hyderabad, India **May 2024**
Bachelor of Technology in Computer Science and Engineering