Pratyusha Adibhatla

pratyushaadibhatla3@gmail.com +1 (682) 374-3740 LinkedIn GitHub

Experience

Research Assistant – Data Scientist, *UCSD Wireless Communication, Sensing, and Networking Lab* *San Deigo, CA*

- Processessing 50+ heterogeneous datasets(HDF5, CSV, DAT, PCD) for ML analysis using Python and SQLite, reducing prep time to under 3 minutes. Enabling downstream modeling and feature extraction for signal and radar data research.
- Automating dataset preparation workflows for research analysis, including metadata extraction, schema validation, and data quality checks, ensuring high-quality inputs for predictive modeling.
- Implementing data lineage and feature retrieval so researchers can query, view, and download datasets with change tracking.

Research Assistant - Data Scientist (Multimodal Data & ML Retrieval), *University at Buffalo* *Buffalo, NY*

- Conducted feature extraction and preprocessing on multimedia datasets (video, image, text) to enable machine learning analysis and semantic retrieval; applied RAG-based embeddings to support research experiments.
- Evaluated multiple embedding strategies using statistical metrics (precision, recall) to optimize low-latency retrieval performance.

Student Assistant – Data Scientist (Analytics & Reporting), *University of Texas at Arlington* *Arlington, TX*

- Designed and implemented a structured SQL database to consolidate fragmented inventory data from multiple facilities, enabling detailed exploratory data analysis and supporting predictive modeling of inventory demand patterns.
- Developed Python-based ETL workflows to clean, transform, and standardize legacy Excel datasets, handling missing values and schema variations, resulting in analysis-ready datasets used for trend identification and forecasting.
- Built a Power BI dashboard to visualize inventory trends, highlighting top 20 most-ordered and bottom 20 least-ordered items, supporting reordering decisions that improved inventory efficiency by 5%.

Programmer Analyst (Data & Application Support), *Cognizant (Client: Liberty Mutual Insurance)* *Coimbatore, India*

- Supported UAT and Production batch workflows processing 100k+ insurance transactions per month, validating claims and payment data to ensure accuracy for downstream analytics and reporting.
- Investigated 15–20 failed batch jobs per week by analyzing upstream data changes, null values, and schema mismatches; identified patterns and escalated systemic issues to development teams.
- Developed and maintained SQL queries to validate claims, payment, and policy datasets, performing source-to-target reconciliation on 200–300 transactions per batch using record counts and financial totals.
- Generated daily and weekly payment analytics reports for 5–10 stakeholders, summarizing failures, trends, failed payment rate and average resolution time, supporting operational and business decision-making.

Programmer Analyst Trainee *Cognizant* *Coimbatore, India*

- Built a Spring Boot (Java) + Angular full-stack application to display bus routes and live location status for public transport users, supporting 5–10 active routes in test and demo environments.
- Designed a relational database with 10+ route fields and REST APIs to handle 100–300 daily location updates, implementing GPS validation and normalization to reduce inconsistent records by 15–20%.

Projects

Award Record Extraction Scraper – UTA Capstone Project

- Built a Python pipeline using RAG and vector embeddings to scrape, clean, and structure award and fellowship records from multiple sources, enabling efficient indexing and retrieval of unstructured data.
- Trained and applied lightweight ML models to classify and extract key fields (e.g., award name, recipient, date) from raw text, improving data accuracy and consistency.

Technologies

Programming & Scripting: Python (Advanced), SQL (Advanced), Java (Intermediate)
Data Analysis & Machine Learning: Python (Pandas, NumPy, Scikit-learn), RAG & Vector Embeddings, Feature Engineering, Statistical Modeling, EDA
Data Visualization & Reporting: Power BI, Matplotlib, Seaborn, Plotly
Databases & Vector Stores: MySQL, SQLite, ChromaDB
Tools & Platforms: Git/GitHub, Docker, REST APIs, Azure Pipelines, ETL/ELT Workflows

Education

Masters in Data Science | *University of Texas at Arlington Arlington, USA*

- Relevant Coursework: Machine Learning, Data Science, Data Analytics, Database Systems, Data Mining, Neural Networks, Probability and Statistics.

BTech in ECE | *Andhra University Visakhapatnam, India*