

TEJODHAY BONAM

bonam.t@northeastern.edu | (+1) 8573950779 | Portfolio: tejodhayb.github.io | [LinkedIn](#)

EDUCATION

Northeastern University - *Khoury College of Computer Sciences*, MS Data Science, Boston, MA May 2024

Coursework: Supervised & Unsupervised Machine Learning, Data Engineering, DBMS, NLP, Algorithms, LLMs

Graduate Teaching Assistant (NLP, Data Science)

- Supported 80+ students in key data science & NLP concepts - (data cleaning, data visualizations, model interpretability, model optimization, hypothesis testing) through scheduled office hours & Python programming assistance.

Sri Sathya Sai Institute of Higher Learning, BS Mathematics & Computer Science, Bengaluru, KA 2021

- Headed the *Western Flute* section of the *Institute Brass Band*; Won *University Badminton Championship* (2 times)

SKILLS & CERTIFICATIONS

Programming Languages/ ML Frameworks/ Libraries: Python, SQL, TensorFlow, PyTorch, Scikit-Learn, Pandas, NLTK

Databases: AWS S3, RDS, RedShift, Azure Synapse Analytics, Azure SQL Database, MySQL, MS SQL Server, MongoDB

Software and Tools: Apache Kafka, Spark, Airflow, Apache Flink, Docker, MS Excel, Power BI, Tableau, Flask, DVC, CI/CD, GitHub Actions, MLflow, Agile, Scrum, Azure Data Factory, Azure Databricks, AWS EMR, SageMaker, Kinesis, Glue, Athena

Certifications: AWS Data Engineer Associate (TBC), Microsoft Azure Data Engineer Associate, AWS Cloud Practitioner

WORK EXPERIENCE

ClearSet.AI

AI/ML Engineer

Jan 2024 – May 2024

Remote

- Collaborated in a fast-paced startup environment on 4+ AI projects - Blog generation LLM app, Enterprise Chat App, Conversational Q&A RAG chatbots leveraging Llama2, OpenAI, Google PaLM APIs to launch product-market prototypes.
- Implemented supervised & unsupervised ML algorithms to analyze 10+ client datasets (structured, semi & unstructured) and synthesized large amounts of data into succinct actionable insights to stakeholders (technical & non-technical).
- Created cool Tableau data visualizations and delivered 15+ Power BI dashboards to identify trends, patterns and anomalies.

Accenture

Data Engineering Associate

Mar 2022 – Sep 2022

Hyderabad, IN

- As part of a top-performing 5-member Agile (Scrum) team, I actively collaborated with data engineering team to deliver 2 projects on time and within budget, applying data engineering best practices to develop robust and scalable data pipelines.
- Developed an E2E real-time data engineering solution to migrate data from on-prem SQL Server database to **ADLS Gen2** leveraging Azure Data Factory, **PySpark** transformations on **Azure Databricks** and loaded into **Azure Synapse Analytics**.
- Automated ETL pipelines orchestrated on **Azure Data Factory** enabling daily refreshes to ingest OLAP data into Azure Synapse Analytics improving data processing efficiency by 35%, transforming data through bronze, silver & gold layers.

Arthashastra Intelligence

Jr. Data Scientist

Jun 2021 - Mar 2022

Hyderabad, IN

Star Projects:

FMCG client - Customer & Purchase analytics using Segmentation, Targeting, Positioning, Marketing Mix, Price Elasticity

- Performed **K-means clustering** combined with **PCA** for customer segmentation. Analyzed purchase behavior through descriptive and predictive models for purchase incidence, brand choice, purchase qty. - (**110% sales growth**)

Appliances Mfg. client - User Centric Forum, integrated with Recommendation engine & a Real-time Streaming Data Pipeline

- Content based (**LDA, tf-idf, NER**), Collaborative filtering (**spectral, k-prototypes**) for personalized articles & targeted Ads. Used **AWS Kinesis, lambda, Glue, Athena** to create live data visualization dashboard (**30% boost in user engagement, CTR**)

PROJECTS

Microsoft - Future Ready Talent

[Facial Emotions based Music Recommendation](#) (OpenCV, Azure Face API, emotion detection, K-Means clustering)

- The web app gets live video feed from web cam, calls Face API that detects emotions. Accordingly, the app will rank the recommended songs employing K-Means clustering. (Often used by my friends and family - **Satisfaction rate: 70%**)

[Fashion Visual Search web app](#): (Azure - Computer Vision, Cognitive Search, Open AI, Speech Services)

- An interactive web app that utilizes vector embeddings to create an efficient and accurate visual search index along with UMAP analysis to search for apparels based on their visual features such as color, pattern, texture. (**Top-N Acc: 92%**)

Portfolio Projects

[Bing News Data Analytics using Microsoft Fabric](#) (Data Factory, One Lake, Synapse - DE, DS, Power BI, Data Activator)

- Configured an E2E data engineering pipeline using Data Factory for ingestion, Synapse DE for transformations along with incremental load, Synapse DS for sentiment analysis of news, Power BI reports & set up Teams alerts using Data Activator.

[Smart City E2E Realtime Data Streaming Pipeline](#) (IOT, Kafka, Apache Spark, Docker, AWS - Glue, Athena, Redshift)

- Designed an E2E data streaming pipeline simulating real-time telemetry data from IOT devices & analyzed in AWS Cloud covering each phase from data ingestion to processing and finally storage.

Northeastern University

[Sports - Football & Olympics Data Analytics](#) (Azure - Data Factory, Data Lake Gen2, Databricks, Synapse Analytics)

- Engineered 2 end-to-end ETL pipelines with similar architecture using Azure Services ingesting data using Airflow & ADF into ADLS Gen2, PySpark transformations on Databricks, analytics using SQL on Synapse, Power BI reports (**Fun projects**)

[PawPal: Petcare Veterinary RAG chatbot](#) (Llama2, LangChain, Multilingual E5 large, FAISS, AWS, FastAPI)

- Built the app on open-source stack useful for Vet Doctors and Pet Lovers to provide rapid (**with a response time of 0.4 sec**) and reliable veterinary advice (with **85% accuracy** of diagnoses and treatment recommendations verified by real Vets).

P.S. - Visit my [portfolio](#) for my recent LLM projects [PawPal](#), [Retail Text2SQL](#), [Playground AI](#) & other [DE](#), DS, NLP projects!