

Harsha Vardhan Reddy Tamatam

• harsha.vardhan@rutgers.edu • +1 617-642-4707 • [linkedin.com/in/harsharshavardhantamatam/](https://www.linkedin.com/in/harsharshavardhantamatam/)

EDUCATION

- RUTGERS UNIVERSITY**, NEW BRUNSWICK, NJ, USA 09/2023 - 04/2025
Master of Science in Data Science, Dept of Statistics
- MAHINDRA ECOLE CENTRALE**, HYDERABAD, INDIA 08/2016 - 05/2020
Bachelor of Technology in Computer Science

WORK EXPERIENCE

- DATA ANALYST INTERN - LS DIRECT**, NY, USA 06/2024 - 12/2024
- Predicted customer relocation using an imbalanced classification model, boosting targeted ad efficiency by 11%.
 - Created vector data stores for 15M+ users, optimizing semantic search retrievals with LangChain & LlamaIndex.
 - Developed a Python module for seamless SQL to BigQuery migration with dbt and Airflow, ensuring data integrity.
- MACHINE LEARNING ENGINEER - VEDANTU LEARNING**, INDIA 11/2021 - 06/2023
- Fine-tuned and deployed a DistilBERT model with PyTorch and FastAPI to intelligently classify and map student queries(or tickets) to tutors in real-time, reducing response times by 65% and boosting chat engagement by 3.5X.
 - Built XGBoost models on student behavioral data to identify high-risk churn users, and developed dashboards to help sales and leadership teams to design targeted retention strategies.
 - Applied DBSCAN & PCA for unbiased clustering, followed by A/B testing and data modelling to evaluate feature impact.
- FOUNDING ENGINEER - KHABRI AUDIO**(YC STARTUP), INDIA 12/2020 - 11/2021
- Trained and deployed a Text-to-Speech platform using Tacotron 2 and WaveNet, with developers dashboard, enabled creators to generate over 600+ hours of native, human-like Hindi audio content.
 - Developed a deep ranking-based content recommendation system, increasing CTR by 19% and purchase rate by 2%.
 - Built data infrastructure from scratch, logging user interactions, storing data, and enabling analytics-driven decisions.
- FOUNDER - AUTODUB**, INDIA 04/2018 - 12/2020
- Founded and led Autodub.in, developing AI-powered speech translation software to enable multilingual accessibility.
 - Scaled AI infrastructure from scratch, enabling rapid video dubbing at 90% lower cost and near-human accuracy.
 - Autodub's technology was acquired by Khabri Audio after national media recognition for dubbing COVID-19 awareness content in Indian languages.

SKILLS

- **Programming:** Python, Pytorch, FastAPI, Scikit-learn, Transformers, LangChain, Flask, Django, Streamlit
- **AI & ML:** Gen-AI, Machine Learning, Deep learning, NLP, LLMs, RAG, ML System Design, Data Modeling
- **Databases & Cloud:** SQL, NoSQL, VectorDB, BigQuery, Redis, Airflow, GCP, AWS, Kubernetes, CI/CD

RESEARCH EXPERIENCE & PROJECTS

- FOUNDATION MODELS EVALUATION & USE-CASE MAPPING** 01/2025 - Present
AI ETHICS LAB, RUTGERS UNIVERSITY Advisor - Dr. Mhamed Bettaieb
- Co-Authoring a research study on existing foundation models, comparing architectures and applications across domains.
 - Developing a structured model selection framework, guiding optimal LLM choices for different real-world scenarios.
- ENTERPRISE TICKET TRIAGE AUTOMATION SYSTEM** Funding - Chase Bank
DATA & STATISTICAL SERVICES LAB, PRINCETON UNIVERSITY
- Fine-tuned MiniLM and CatBoost models to autonomously route support tickets to the correct teams and predict priority levels, achieving over 90% classification accuracy.
 - Built a real-time FastAPI backend and a lightweight Streamlit dashboard for Support Ops to monitor, search, and override ticket triage decisions.
- CSHORTS - AI NEWS SUMMARIZATION PLATFORM - DEMO**
- Fine-tuned and quantized a T5 model for abstractive summarization, reducing model size by 45% and optimizing for low-cost, scalable inference; deployed the model on AWS SageMaker.
 - Engineered a FastAPI backend to ingest, process, and store news articles from APIs and web scrapers into a NoSQL database; developed a Streamlit app to deliver categorized, real-time news summaries within 60 seconds.