RAHUL NENAVATH

nenavath.r@northeastern.edu | +1 857-763-7170 | [LinkedIn](#) | [GitHub](#) | [Portfolio](#) | Boston, MA

Data Scientist with **3 years of experience** in leveraging data to support businesses in automated and strategic decision-making. Extensive expertise in Generative AI, Natural Language Processing, and Analytics domains.

EDUCATION

Northeastern University   Boston, MA
Master of Science in Data Analytics Engineering   Expected May 2025
Coursework: Data Mining, Statistical Methods, Statistical Learning, Machine Learning Ops, Data Visualization

Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram   Chennai, India
Dual Degree (B.Tech and M.Tech) in Computer Engineering   Jul 2016 - May 2021
Coursework: Database Systems (DBMS), Algorithms, Analytics & Systems of Big Data, Digital Image Processing

---

WORK EXPERIENCE

Data Science Intern   May 2024 - Aug 2024
Asurion LLC   Nashville, TN
- Built a **GenAI** autocomplete feature for call center agent application, **saving $575,000 annually** in labor cost.
- Designed model data pipeline to utilize real-time call transcripts, call summaries, and messaging history to autocomplete messages typed in the **call center chat application**.
- Developed Retrieval Augmented Generation (**RAG**) pipeline with Qdrant vector DB to improve autocomplete quality.
- Designed **LLM Judge** framework using 7 subjective attributes (Relevancy, etc) to measure autocomplete quality.
- Benchmarked **GPT-3.5**, **GPT4**, & **Claude Haiku** on **Azure OpenAI** & **AWS Bedrock**.

Graduate Research Assistant   Mar 2024 - Current
Northeastern University   Boston, MA
- Built **BM25** & **HNSW indices** to enable human rights researchers to search **over millions** of **supply chain records**.
- **Fine-tuned** sentence transformer model on English & Chinese company names to **improve retrieval accuracy**.
- Utilized **Dask** to **parallel process** the workloads and compute batch search results.

Data Science Engineer   Aug 2021 - Jul 2023
Twimbit Research India Pvt Ltd   Chandigarh, India
- **Enhanced search** engine **accuracy by 8x** by redesigning the search engine using vector DB and LLM reranker.
- **Fine-tuned** a sentence-transformer model on 1,500 in-domain question and answer pairs generated from **ChatGPT**.
- Built an **event processing pipeline** using Redis and PostgreSQL to process and store **10K daily user events.**
- Designed **15 SQL analytics queries** for user and expert dashboards, tracking vital content consumption metrics.
- **Optimized** query response time by **64%** by updating the record structure in the **Algolia Search Index**.

Artificial Intelligence (AI) Engineer Intern   May 2020 - Jun 2021
CloudSEK Information Security Pvt Ltd   Bangalore, India
- **Improved** phishing domain-name detection module **accuracy by 40%** by developing Character-based Siamese LSTM, GRU, and 1-D CNN Models in **TensorFlow**.
- Trained the models on a domain name dataset of 3,500 samples and validated the models using Adversarial Testing.

Software Engineer Intern   Jun 2019 - Jul 2019
Hexagon Capability Center Pvt Ltd   Hyderabad, India
- Developed a chatbot as a first response to Human Resource (HR) inquiries using **Google's DialogFlow**.

PUBLICATIONS

Subin Sahayam, **Rahul Nenavath**, Umarani Jayaraman, Surya Prakash - **"Brain tumor segmentation using a hybrid multi-resolution U-Net with residual dual attention and deep supervision on MR images"** - [paper](#)
Biomedical Signal Processing and Control, Elsevier, Volume 78, pp. 103939, September 2022

PROJECTS

Keyword Generation by Fine-tuning BART Transformer Model [GitHub](#)   Aug 2022
- Fine-tuned BART model on open-source keyword datasets, achieving a 33% Rouge score for keyword generation.

SKILLS | TECHNOLOGIES

Generative AI: OpenAI & Anthropic APIs, Few-shot & Chain-of-Thought prompting, LLM evaluation, Synthetic Data Gen
Data Engineering: PostgreSQL, Redis, DuckDB, MongoDB, Airflow, Pinecone DB, Weaviate DB, Qdrant DB
Data Science & Machine Learning: Pandas, Numpy, Scikit-Learn, Statsmodels, Spacy 3, Tensorflow, Pytorch
Data Visualisation: Tableau, PostHog, Apache Superset
Deployment: FastAPI, Docker Compose, CI/CD - GitHub Actions, MLFlow, DVC
AWS Stack: SageMaker, Bedrock, S3, EC2, Lambda, SQS, ECR, Secrets Manager