

# Anurag Sanjay Ghosh

Boston, MA | (857)-540-6982 | [ghosh.anu@northeastern.edu](mailto:ghosh.anu@northeastern.edu) | [LinkedIn](#) | [GitHub](#)

## WORK EXPERIENCE

---

**Ribbon Communications**, Westford, MA

January 2025 – Present

### Data Scientist

- Built an end-to-end log analysis platform to ingest raw, unstructured logs and transform them into feature-enriched, clustered streams of correlated events. Leveraged **BERTopic** modeling for dynamic grouping of semantically related log events.
- Conducted an extensive literature review to integrate advanced processing techniques, including a novel algorithm for automatic log parsing to extract event templates and entities, and unsupervised methods for grouping temporally correlated events.
- Secured a significant purchase order from **Altice**, expanding enterprise adoption of Ribbon's Log analytics capabilities.
- Engineered a ML pipeline for optical performance monitoring, leveraging in-band power signal features to train quantile XGBoost and Gaussian Process regressors for **Optical Signal-to-Noise Ratio (OSNR)** estimation.
- Applied **physics-informed** feature engineering to mitigate data sparsity effects with a **twofold** increase in predictive fidelity.
- Built a chat-driven analytics platform for conversational EDA, automating chart/dashboard creation in **Apache Superset** using a multi-agent LlamaIndex workflow with a **self-correcting SQL** agent and **MCP-exposed** Superset APIs as agent tools.

### Data Science Intern

- Developed a **website categorization** service to catalog & capture network traffic trends with a **RoBERTa** model trained on **1M** webpages scraped by **Selenium** crawlers. Added subcategorization with **Few-Shot DeBERTa** model for additional enrichments.
- Delivered an F1 score of **85%** & reduced unknown/new URL cataloging & processing time by **70%** on client side (**AT&T**).

**Universal Music Group**, New York City, NY

July 2023 – December 2023

### Data Science Analyst Intern

- Trained **LightGBM** based **Customer LTV Models** on **Sagemaker** for predicting purchase propensity & to identify superfans (high value customers). Processed **4 million online transactions** to train the model, & interpreted predictions using **Shapley**.
- Elevated leadership's awareness of a **data-driven** approach to fanbase analysis & customer valuation, leading to its prioritization.
- Prepared a **feature extraction module** for promotional emails to identify engagement driving features.
- Devised a **BERT-based active-learning method** with unsupervised clustering to predict email categories & recommend key linguistic & structural attributes contributing to high customer engagement rates.
- Secured a **10% increase** in performance of downstream **in-production** engagement assessment models.

**WINES Lab Northeastern University**, Boston, MA

January 2023 – September 2023

### Machine Learning Research Assistant

- Explored machine learning **based radio fingerprinting** by training 2 deep learning models, **Alex net & Vision Transformers** on sequences of in-phase (I) and quadrature (Q) data from **LoRa** devices to identify the devices in varying external scenarios.
- Attained **99%** classification accuracy across all device configurations under same-day training and testing conditions.

**TIAA GBS**, Mumbai, Maharashtra

July 2019 - July 2022

### Analyst, Software Developer

- Led a team of 4 developers overseeing **development & production** of **multiple** spring boot microservice applications.
- **Independently** supported & collaborated with the business on live deployments for **retirement income evaluation apps**.
- Oversaw **20** successful **production releases** with **OpenShift** integration reducing deployment time from hours to a few minutes.

## ACADEMIC PROJECTS

---

### Store Sales Prediction on Time series Data

February 2023 - April 2023

- Programmed **LSTM** and **XGBoost** regression models to predict sales of a store from previous time series data of sales.
- Performed **EDA** to find the periodicity & correlation of sales with factors like oil prices, holiday events and promotions.
- Achieved R-squared value of **0.98** on training and **0.82** on testing data, with XGBoost and LSTM models respectively.

## PUBLICATIONS

---

“Hybrid Image Encryption Technique Using Genetic Algorithm and Lorenz Chaotic System” ITM Web of Conferences 32, 03009 (2020), [https://www.itm-conferences.org/articles/itmconf/pdf/2020/02/itmconf\\_icacc2020\\_03009.pdf](https://www.itm-conferences.org/articles/itmconf/pdf/2020/02/itmconf_icacc2020_03009.pdf)

## EDUCATION

---

**Northeastern University, Khoury College of Computer Sciences**, Boston, MA

December 2024

Master of Science, Artificial Intelligence

4.00 GPA

Relevant Coursework: Large Language Models, Reinforcement learning, NLP, Algorithms, Machine Learning.

**Teaching Assistant:** Foundation of Artificial Intelligence (CS5100)

**Mumbai University**, Mumbai, India

June 2019

Bachelor of Engineering, Computer Engineering

## TECHNICAL SKILLS

---

**Programming/Scripting:** Python, JavaScript, Java, C++, C.

**Databases and Cloud Service:** Google Cloud Platform, Big Query, AWS, Sagemaker. Azure, MySQL, Microsoft SQL Server

**Tools & Frameworks:** pandas, Docker, Python (Keras, PyTorch, TensorFlow, Pandas, Scikit-learn, Streamlit, Llamaindex, LangGraph)