Raghu Ram Sattanapalle

Boston, MA, USA - 02119 • +1 (347) 873-2177 • sattanapalle.r@northeastern.edu
linkedin.com/raghuramsattanapalle • github.com/RaghuRamSatt • raghuramsatt.github.io

Education

Northeastern University — Expected May 2025
Master of Science in Data Science (GPA: 4.00 / 4.00) — *Boston, MA, USA*
- **Relevant Coursework:** Large-Scale Parallel Data Processing, Supervised Machine Learning, Database Management

New York University — May 2018
Master of Science in Mechanical Engineering — *New York, NY, USA*
- **Relevant Coursework:** Robot Perception, Simulation Tools and Software for Mechatronics and Robotics

Technical Skills

Programming Languages: Python, R, MATLAB, C++, SQL, Scala, Julia, JavaScript, HTML, CSS
Machine Learning: TensorFlow, PyTorch, Keras, FastAI, Scikit-learn, Deep Learning, Time Series Analysis, Pattern Recognition, Ensemble Methods, Natural Language Processing (NLP)
Data & Cloud Technologies: AWS, Hadoop, Spark, Hive, HBase, PostgreSQL, MySQL, MongoDB
Data Visualization: Tableau, Power BI, Plotly, Matplotlib, Seaborn, Bokeh, ggplot2, D3.js, Excel
Software Development: Git, Docker, CI/CD

Experience

Northeastern University, Khoury College of Computer Sciences — Sept 2023 - Dec 2023; July 2024 - Present
Graduate Teaching Assistant - CS5800 Algorithms — *Boston, MA*
- Facilitated graduate-level algorithms education for over 35 students each semester, covering topics from fundamental data structures to advanced algorithms, including Dynamic Programming, Graph Algorithms, and NP-complete problems.
- Provided comprehensive support through grading assignments, offering constructive feedback, and conducting weekly office hours, enhancing student understanding and performance.

Veeco Instruments — Jan 2024 - June 2024
Engineering Data Scientist (Co-op) — *San Jose, CA*
- Developed initial convolutional neural networks (CNNs) using TensorFlow and PyTorch to predict boron wafer resistance (RS) for semiconductor manufacturing equipment, with an initial average error rate of 13-22%.
- Implemented domain-specific data augmentation and group normalization techniques, significantly improving the model's accuracy and ultimately reducing the average error rate to 6%.
- Automated data extraction and structuring from storage drives, achieving 100% data capture, highlighting skills in efficient data pipeline creation for large-scale analysis.
- Developed custom visualization tools for manufacturing equipment channel data, facilitating the rapid identification of trends and anomalies, which enhanced process control and informed decision-making in semiconductor tool fabrication.

NYU Dynamical Systems Laboratory — Sept 2018 - Jan 2019; June 2019 - Aug 2021; Jan 2022 - Aug 2022
Researcher/ Research Assistant — *Brooklyn, NY*
- Spearheaded the development of machine learning models using the MIMIC IV dataset (300M+ clinical observations) to predict ICU patient mortality rates, achieving 90% accuracy and contributing to improved patient care.
- Conducted a causal inference study on gun prevalence and mass shootings using time series analysis (TRAMO-SEATS) and transfer entropy, resulting in a publication in Nature Human Behaviour and securing $2.1M in NSF funding.
- Engineered innovative stochastic differential equation models using information theory for zebrafish behavior, advancing understanding of collective systems and leading to a publication in Flow: Applications of Fluid Mechanics.

Projects

Optiver - Trading at the Close: Predict US Stock Movements | *Python, LightGBM, XGBoost, PyTorch*
- Developed quantitative machine learning trading models to predict stock price movements during the closing auction for NASDAQ-listed stocks, achieving a top 20% ranking in a Kaggle competition.
- Preprocessed 5M+ data points, engineered features, and applied advanced statistical techniques like feature scaling, temporal lagged variables, and rolling window statistics to optimize model performance.
- Implemented and fine-tuned LightGBM, XGBoost, and Neural Networks, focusing on minimizing Mean Absolute Error (MAE) for accurate stock price prediction.

Soundit: Database-Driven Music Streaming Platform | *MySQL, JavaScript, Python, Recommendation Systems*
- Architected a music streaming platform with MySQL database supporting 1M+ records for user interactions, playlists, and music metadata.
- Developed a recommendation system using collaborative filtering and content-based approaches, enhancing user engagement and music discovery.
- Engineered features including dynamic playlist generation, social sharing, and analytics tracking user listening patterns.

Environmental Impact Analysis: Air Quality During COVID-19 | *Python, SQL, Data Visualization*
- Led a team investigation correlating COVID-19 lockdown measures with air quality in Massachusetts using the Air Quality Index (AQI) and Social Distancing Index (SDI).
- Processed and analyzed large datasets from multiple sources, creating insightful visualizations to communicate findings effectively.
- Developed a comprehensive analysis framework, applicable to similar environmental impact studies in urban areas.