Raghu Ram Sattanapalle

Boston, MA +1 (347) 873-2177 sattanapalle.r@northeastern.edu
linkedin.com/in/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*
- **Coursework:** Large-Scale Parallel Data Processing, Supervised Machine Learning, Database Management, Algorithms

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

Experience

Northeastern University, Khoury College of Computer Sciences Sept 2023 - Dec 2023; July 2024 - Present
Head Graduate Teaching Assistant - CS5800 Algorithms *Boston, MA*
- Lead the instruction of complex algorithmic concepts to over 40 graduate students, covering topics such as **Dynamic Programming, Graph Algorithms, and NP-complete problems**, enhancing their analytical and problem-solving skills.
- Manage and coordinate a team of three TAs, delegating responsibilities for grading, code demonstrations, and online office hours, while personally overseeing in-person office hours and lecture recitations.

Veeco Instruments Jan 2024 - June 2024
Engineering Data Scientist (Co-op) *San Jose, CA*
- Optimized semiconductor manufacturing systems by developing **convolutional neural networks (CNNs)** using **Python, TensorFlow, and PyTorch** to predict boron wafer resistance, reducing error rates by up to **73%**.
- Automated **data extraction and processing pipelines** using **Python and SQL**, achieving **100% data capture** from storage drives, streamlining data accessibility and enhancing analysis efficiency across semiconductor manufacturing processes.
- Created automated custom **visualization tools** to identify trends and anomalies in manufacturing data, facilitating **data-driven decision-making** and enhancing process control.

NYU Dynamical Systems Laboratory Sept 2018 - Jan 2019; June 2019 - Aug 2021; Jan 2022 - Aug 2022
Researcher/ Research Assistant *Brooklyn, NY*
- Secured a **$2.1M NSF grant** as co-author to investigate the U.S. firearm ecosystem, applying **data science and network theory** to analyze firearm prevalence, legislation, and socioeconomic factors.
- Collaborated with **NYU Langone medical experts** to develop **machine learning models** on the **MIMIC dataset (300M+ clinical observations)**, achieving **90% accuracy** in predicting ICU patient mortality rates and enhancing clinical efficacy and safety.
- Led a **causal inference study** on mass shootings, media coverage, and firearm acquisition in the U.S., using **R and Python** for **time series analysis and transfer entropy**, leading to a publication in *Nature Human Behaviour*.
- Modeled zebrafish behavior using **stochastic differential equations** in **MATLAB and Mathematica**, contributing to advancements in understanding collective systems through **statistical analysis and interdisciplinary collaboration**.

Projects

Scalable Music Similarity Analysis with Spark Oct 2024 - Dec 2024
- Implemented distributed **K-Means** clustering in **Spark** to analyze music patterns, and a novel **H-V partitioning** strategy for **collaborative filtering** to compute song-to-song similarities, both applied to the **Million Song Dataset**.
- Achieved a **4.58x speedup** for K-Means and demonstrated scalability with a **5x larger dataset** for similarity computations; performed on **AWS EMR** clusters with up to **8 nodes** (**7 workers**).
- Optimized **Spark** performance by refining data partitioning, utilizing caching and broadcasting techniques, and mitigating driver-side bottlenecks, as confirmed through **Spark UI** analysis.

Optiver - Trading at the Close: Predict US Stock Movements Oct 2023 - Dec 2023
- Collaborated with a team of three to develop quantitative **machine learning (ML)** trading models aimed at predicting stock price movements during the closing auction for NASDAQ-listed stocks, achieving a top 20% ranking in a Kaggle competition.
- Processed over 5 million data points, engineered features, and applied advanced statistical techniques like **feature scaling, temporal lagged variables, and rolling window statistics** to optimize model performance.
- Utilized 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 Oct 2023 - Dec 2023
- Designed a scalable **MySQL** database schema managing a complex structure of user data, playlists, artists, albums, tracks, and **1M+ songs**, demonstrating proficiency in relational database design.
- Developed a **Python Flask** backend with **RESTful APIs** for authentication, music playback, playlist management, and simulated-recommendations; integrated with a **Vue.js** frontend using **Node.js** and **npm**.
- Engineered recommendation features using **collaborative** and **content-based filtering** methods; implemented robust error handling/logging.

Technical Skills

Programming Languages: Python, R, MATLAB, C++, Java, Scala, Julia, JavaScript, HTML, CSS, SQL
Machine Learning: TensorFlow, PyTorch, Keras, Scikit-learn, Deep Learning, Time Series Analysis, Ensemble Methods, NLP
Data Engineering & Cloud Technologies: AWS (EMR, EC2, S3), Hadoop, Spark, BigQuery, Hive, dbt
Database/Data Stores: Relational Database Design, SQL, MySQL, PostgreSQL, MongoDB (NoSQL)
Data Visualization: Tableau, Power BI, Looker, Matplotlib, Seaborn, Plotly, Bokeh, ggplot2
Software Development/OS: Git, Docker, CI/CD, Agile Methodologies, Bash, Kubernetes, **Linux**