Giri Manohar Vemula

Boston, MA | vemula.gi@northeastern.edu | 857-832-1320 | LinkedIn | GitHub | Portfolio

EDUCATION

Northeastern University, Boston **Dec 2024**
Master of Science in Data Analytical Engineering
Coursework*:* Database Management, Computation and Visualization, Statistical Learning, Applied NLP

Vellore Institute of Technology, India **Jun 2022**
Bachelor of Technology in Computer Science and Engineering
Coursework*:* R and Python Programming, Machine Learning, Optimization Techniques, Artificial Intelligence

SKILLS

Programming Languages	**:** Python (Pandas, Tensor-flow, Seaborn), R (tidyr, readr, caret), Java
Data Engineer Tools	**:** SQL (MySQL, PostgreSQL), NoSQL (MongoDB, Neo4j), AWS, Snowflake
Data Visualization Tools	**:** Tableau, Flourish, Data Wrapper, Advanced Excel, Power BI
Specializations	**:** ETL, Data Analysis, Data Modelling, Data Mining, Statistics, Data Wrangling
Soft Skills	**:** Story Telling, Problem-Solving, Collaboration, Critical Thinking

PROFESSIONAL EXPERIENCE

Massachusetts General Hospital, Boston, MA **Feb 2024 – Present**
Data Science Research Intern

- Leading a maternal mental health research project at Dekel Lab, engaging in data analysis, machine learning, statistical analysis, literature review, and manuscript preparation for scientific publication
- Providing analytical support for additional lab projects, with a focus on healthcare data concerning women's pregnancy and post trauma
- Developing scalable, efficient modeling algorithms and data solutions for handling large-scale data, optimizing processing and analysis
- Developing self-service tools and dashboards to facilitate exploratory research and patient segmentation

ACADEMIC PROJECTS

Covid Analysis using Amazon Web Services [AWS S3, Glue, Redshift, Athena] **Jan 2024**

- Constructed a scalable and efficient data warehouse using AWS Redshift, capable of handling a dataset with over 10,000 entries for COVID-19 analysis to ensure secure data storage on Amazon S3
- Executed a Glue crawler for data structuring, and carried out ETL processes to enhance data quality
- Employed Athena for querying CSV datasets and migrated the necessary data to Redshift that streamlined analytical processes, enabling a deeper analysis of COVID-19 data trends

Personality Prediction Test, [Python, Supervised ML, Data Cleaning, Tableau] **Apr 2023**

- Predicted personality types using machine learning techniques based on a variety of input data by selecting and training various machine learning models to find the most accurate personality
- Implemented diverse models such as Decision Trees, Neural Networks, KNN, logistic regression, and PCA. Also performed error analysis, lift charts, and ROC curves to evaluate model performance
- Efforts resulted in development of a robust personality prediction system, additionally levered Tableau dashboard that provided clear visualizations of participant demographics and behaviors

Human Activity Recognition, [Time Series Complexity Analysis, R] **Dec 2022**

- Analyzed physical activity patterns using time series data from 15 participants to prevent fall incidents, performing a detailed time series complexity analysis on activities like walking, running, and climbing
- Applied NVG and HVG algorithms to time series data; created scatter plots, and computed network topology metrics like average degree, network width, and path length
- Identified patterns and connections that potentially anticipate falls and avoid human activity-related issues

CERTIFICATES/ CO-CURRICULAR ACTIVITIES

Data Mining Hackathon (US Offense Type), Special Recognition in Hackathon Winners
Office Bearer and Poster Designer, Robotics Club, VIT University
Certificate of Merit, VIT-AP Engineering Clinics Projects