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 : R (tidyr, readr, caret), Python (Pandas, Tensor-flow, Seaborn), Java
Data Visualization Tools : Tableau, Flourish, Data Wrapper, MS Excel, Power-BI
Data Engineer Tools : SQL (MySQL, PostgreSQL), NoSQL (MongoDB, Neo4j), AWS, Snowflake
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**

- Developed data warehouse on AWS Redshift by integrating dataset of 10,000 entries
- Utilized Amazon S3 for data storage and executed a Glue crawler for data structuring and performed ETL processes to refine data quality and ensure efficient usability
- Employed Athena for querying CSV datasets and seamlessly migrated pertinent data to the Redshift warehouse, facilitating comprehensive analytical processes

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

- Explored personality prediction through machine learning, training, and testing diverse models like Decision Tree, Neural Network, KNN, logistic regression, and PCA to identify the most effective approach
- Ensured data quality and accuracy by cleaning unstructured data, enabled precise analysis,conducted comprehensive. Error Analysis, Lift Charts, and ROC curves for model accuracy and performance
- Leveraged Tableau to create highly interactive dashboards, unveiling valuable insights on participant demographics and behavior, providing concise visualizations for analysis

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

- Conducted time series analysis on 15 participants data, focusing on activities like walking, running, and climbing to predict fall incidents using NVG and Horizontal Visibility Graph HVG methods
- Created scatter plot visualizations and computed essential network topology metrics, including average degree, network diameter, and path length, to uncover potential patterns or correlations within the data

CERTIFICATES/ EXTRACURRICULAR ACTIVITIES

Data Mining Hackathon (US Offense Type), Special Recognition in Hackathon Winners
Certificate of Merit, VIT-AP Engineering Clinics Projects