GIRI MANOHAR VEMULA

Boston, MA, 02115 | vemula.gi@northeastern.edu | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

Education

Northeastern University, Boston	Sep 2022 - Dec 2024
Master of Science in Data Analytical Engineering	CGPA: 3.7/4
Vellore Institute of Technology, India	Aug 2018 - Jun 2022
Bachelor of Technology in Computer Science and Engineering	CGPA: 3.6/4

Skills/Knowledge

Programming Languages :	SQL (DbGate), Python (SciPy, SQLite 3, ggplot2), R (tidyr, caret), Java
Data Engineer Tools :	AWS, MySQL (Spark, SSAS), NoSQL (MongoDB, Elasticsearch), Snowflake
Data Visualization Tools :	Tableau CRM, Flourish, Data Wrapper, Advanced Excel, Power BI, Looker
Expertise :	ETL, Web Scraping, Probability and Statistics, Git, Office Suite
Soft Skills :	Story Telling, Interpersonal Skills, Problem Solving, Strategic Thinking
Honors :	Hackathon Winners (Special Recognition), Office Bearer and Project Assistant
Certificates :	Certificate of Merit (VIT), AWS Certified Machine Learning (MLS-C01)

Professional Experience

Data Science Research | Massachusetts General Hospital, Boston Feb 2024 – Jun 2024

- Implemented NLP for symptom tracking of critical maternal medical insights, including Cesarean and preterm birth, using SHAP analysis to reduce irrelevant data by 36% resulting in decrease false positives
- Collaborated with over three medical and computational teams to design A/B tests of developed ML models and with evaluation and testing of models identifying key biomarkers and patterns in PTSD patients
- Utilized Excel and IBM SPSS for statistical analysis and built predictive modeling techniques that achieved early diagnosis of potential patients, reducing 43% time for treatment initiation
- Literature review and presentation under Harvard for manuscript publication of Stillbirth, PCL-5 PTSD, others

Data Analyst | Shiash Info Solutions, India Jan 2022 – Jun 2022

- Oversaw full lifecycle of data processing and production, ensuring regulatory standards, and led initiatives using advanced analytics that identified inefficiencies and saved over 50 K in disputable claims
- Crafted impactful 12 KPI dashboards to executive team and helpful in development of PowerBI and Excel macros to improve on historical and current flase claims, identifying trends and inefficiencies
- Point of contact to conduct user research through surveys and site visits to identify common challenges
- Facilitated meetings for business decisions and liaised between IT teams to rectify data quality issues

Academic Projects

Finance Tone and Reporting BI Tool | NLTK VADER, Hugging Face, MLP Regressor, flashtext

- Web scraping and NLP tool to analyze financial news for business intelligence, leveraging APIs like GoogleNews to track and summarize daily news sentiment about companies such as Amazon

Stock Market Outlook | fredapi, Statsmodels, GCP (Vertex AI, Big Query), GitHub Actions

- Conducted 58 py tests to identify errors and perform coverage report and alerts of deploy analytical workflows
- Deployed end to end automated model monitoring framework and retraining 0.75 threshold accuracy protocols and address model decay or data drift, including logging to track model post-deployment

ECommerce Dive Deep Sales Analysis | Airflow, AWS, Alerts, Power BI

- Performed feature engineering with PySpark, transforming 100k+ retail transactions into finalized features, reducing data exploration steps by 40% and runtime by 27%

Personality Prediction Test | plotly, EDA, clustering, Anomaly detection, Supervised ML

- Built machine learning models from scratch to enhance performance, tested on a 32% redundant complex dataset through strategic data cleaning and partitioning