Lohit Marla

+1 8602347561 | lohitmarla@gmail.com | GitHub | LinkedIn | Portfolio | United States

Education

Master's in Data Science**,** 3.8 / 4.0 GPA**, University of Connecticut** December 2024

Bachelor of Technology, Computer Science and Engineering, 8.8 / 10 GPA, **GITAM University** August 2020

Skills

Data Analysis & Visualization: Pandas, NumPy, Scikit-learn, Seaborn, PyTorch, TensorFlow, Power BI, A/B Testing, Excel
Modeling: Machine Learning, Statistical Modeling, Data Mining, Predictive models, Generative AI Fundamentals, Data Ethics
Programming and Data Engineering: R, Python, OOPs, SQL, Pyspark, Airflow, ETL, AWS, GitHub, Terraform, Snowflake
Certifications: AWS CCP, SQL, Databricks Generative AI Fundamentals, OCI Generative AI Professional**,** Apache Airflow

Experiences

ML Engineer Intern - Nokia, Parker Dewey October 2024

- Designed an end-to-end object detection model workflow, covering dataset preparation, feature engineering, model selection, and real-time object detection using Python.
- Developed a project framework for real-time object detection using pre-trained model such as MobileNetV2 for small dataset and custom model for custom datasets, ensuring comprehensive model evaluation and optimization.

Data Engineer, Parker Dewey August 2024 – October 2024

- Architected and deployed an end-to-end data pipeline to gather data from 8 government sources via APIs using Microsoft Power Connectors, ensuring seamless integration between the data sources and the PowerApps.
- Updated 15 tables in Dataverse and integrated with Power BI, automation of data pipelines to provide low latency insights and improve data accessibility for decision-making, increasing data flow and reporting accuracy by 50%.

Business Intelligence Intern, HexStream June 2024 – August 2024

- Conducted extensive data cleansing, addressing gaps and inconsistencies in raw datasets through imputation by performing preliminary analysis, which enhanced the accuracy and reliability of the reports of forecasts by 35%.
- Generated 10 interactive KPI dashboards based on the business knowledge gained, forecasting dashboards using power query, DAX and demonstrated to non-technical stakeholders for better risk management.

Data Scientist, Parker Dewey May 2024 - June 2024

- Led a detailed competitive data analysis on 25 job postings, strategized data-driven recommendations, and increased application counts by 15%, enhancing team performance.
- Synthesized 15 actionable insights from complex datasets in Excel by integrating, sanitizing and visualizing data accurately with R and Python, driving strategic decisions by statistical methodology with keen attention to detail.

Data Engineer, Novus Platform Sep 2019 - July 2023

- Automated processes of data validation systems by distributed frameworks like Scala Spark for data quality controls, data profiling checks, presented the metrics by Tableau reporting tool to the technical audience.
- Collaborated with cross-functional teams with effective communication via Jira for requirements gathering and developed CI/CD pipelines using GitHub to ensure high-quality code, deployments cycle using Jenkins.
- Conducted RCA, debugging, and troubleshooting production issues by executing and tuning complex SQL queries, resulting in improved query performance and a 55% reduction in operational costs.
- Enhanced data security, HIPPA compliance by implementing data masking and profiling for 1M+ PHI/PII of healthcare data using Batch, Docker, Glue, improving data accuracy by 25% and guided team members.

Capstone Projects

Data Scientist, Indeed

- Conducted Exploratory Data Analysis using Python on a 3.5M-row job posting dataset, employing data enrichment, imputation to improve sponsor satisfaction and demonstrated to the business team for the decision making.
- Implemented various machine learning predictive models, including Linear Regression, LGBM, XGBoost, gradient boosting, and neural networks, achieving a prediction accuracy of 50%, MAE of 1%, and RMSE of 29%.

Forecasting Freight Total Index Value, University of Connecticut

- Developed ARIMA, ARIMAX, and GARCH models using R to forecast Total Service Index values with over 85%.
- Analyzed 288 monthly data points from 2000-2023, achieving 92% explanatory power in identifying patterns and trends in transportation service demand.

Achievements

- Finalist in the 2024 BI & LDP Case Competition among 400 students, presented an innovative AI architecture using AWS for the health insurance, emphasizing efficient data extraction, ingestion, and analytical processing strategies.
- Secured two bronze medals in World Sprint 13 Data Structures Coding Competitions organized by Hacker Rank.